Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Probability Learning in Children: H. W. STEVENSON AND E. F. ZIGLER 185
The Interaction of Ability and Amount of Practice With Stimulus and Response Meaningfulness (*m*, *m'*) in Paired-Associate Learning: V. J. CIEUTAT, F. E. STOCKWELL, AND C. E. NOBLE 193
Some Variables Affecting Latent Extinction: A. R. THOMAS 203
Studies of Distributed Practice: XVIII. The Influence of Meaningfulness and Intralist Similarity of Serial Nonsense Lists: B. J. UNDERWOOD AND J. RICHARDSON 213
An Investigation of Mediation in Preconditioning: R. J. SEIDEL 220
A Comparison Between Rote Learning and Concept Formation: R. METZGER 226
The Interaction of Pitch and Loudness Discriminations: J. D. HARRIS, A. G. PIKLER, H. S. HOFFMAN, AND R. H. ENMER 232
The Effects of Response Contingent Probabilities Which Favor Response Change: M. W. SHELLY, II 239
Stimulus Spacing and the Judgment of Loudness: J. C. STEVENS 246
Resistance to Extinction as a Function of Continuous or Intermittent Presentation of a Training Cue: M. H. MARX 251
The Effect of Intertrial Responding on Conditioning and Extinction of Avoidance Behavior: J. V. MURPHY AND R. E. MILLER 256
The Recognition, Naming, and Reconstruction of Visual Figures as a Function of Contour Redundancy: N. S. ANDERSON AND J. A. LEONARD 262
Body Temperature and Psychological Ratings During Sleep Deprivation: E. J. MURRAY, H. L. WILLIAMS, AND A. LUBIN 271
Deprivation and Generalization: W. O. JENKINS, G. R. PASCAL, AND R. W. WALKER, JR. 274
Gradients in Response Percentages as Indices of Nonspatial Generalization: B. BASS 278
Effects of Previous Experience and Information on Performance on a Word-Formation Problem: W. F. BATTIG 282
Time Sharing as an Index of Automatization: H. P. BAHRICK AND C. SHELLY 288
The Temporal Factor in Reward and Nonreward of Monkeys: R. W. LEARY 294

American Psychological Association
Vol. 56, No. 3 September 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. and R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 56, NO. 3 SEPTEMBER, 1958

PROBABILITY LEARNING IN CHILDREN [1,2]

HAROLD W. STEVENSON AND EDWARD F. ZIGLER

University of Texas

There has been a great deal of interest during recent years in the investigation of probability learning in human *Ss*. In a large number of studies (**1, 4, 5, 10, 11, 12, 14**) it has been found that *Ss* tend to match the stimulus probabilities with their guesses. At the same time, it has been found that under certain conditions *Ss* tend to maximize their frequency of choosing the more frequently reinforced stimulus. Among these are the presentation of a gambling task or a game of skill rather than a problem-solving task (**7, 9**), an increased number of training trials (**3, 6**), the presence of a distracting stimulus (**18, 19**), and the lack of information concerning the appropriateness of some responses (**2, 3, 14**).

These studies have been done primarily with college students as *Ss*. It would be of interest to determine whether behavior similar to that of these *Ss* might be found with other populations of *Ss*, and to explore other conditions which might influence the tendency of *Ss* to choose the more frequently reinforcing stimulus. The present study, therefore, is concerned with testing normal and institutionalized feebleminded children in a situation similar to that which has been presented to adult *Ss*, and with determining whether the behavior of normal children in a probability learning task can be influenced by pretraining experience with different probabilities of reinforcement. Three experiments are presented and the hypotheses tested in each are discussed in later sections.

EXPERIMENT I

The purpose of this experiment is to determine whether the responses of normal children in a probability learning task are similar to those of adult *Ss*. A three-choice, contingent procedure was used in which the response to only one stimulus was reinforced. In this situation, adult *Ss* have been found (**14**) to show a rapid increase in frequency of choice of the reinforced stimulus and to reach an asymptotic level of response slightly

[1] The writers wish to express their appreciation to Morton Weir for testing the *Ss* and for suggestions concerning the conduct of the study. They also wish to thank Raymond Vowell, Superintendent, and Barbara Nellis, Chief Psychologist, of the Austin State School, and the Directors of the Tarrytown Methodist, Jack and Jill, Donnelly, and St. Austin Nursery Schools.

[2] This study was supported by a grant from the National Science Foundation (Grant 3280) to the senior author.

above that corresponding to the frequency of reinforcement.

Method

Subjects.—The *S*s consisted of 45 children attending nursery schools in Austin, Texas. The *S*s were selected at random from among children of the appropriate CA enrolled in the schools. The average CA was 5.5 yr.

Apparatus.—The apparatus consisted of a yellow vertical panel 22 in. long and 16 in. high on which was centered a row of three identical black knobs. The knobs, 1¼ in. in diameter, projected 1 in. from the panel and were 2 in. apart. A red signal light was mounted on the midline, 2 in. from the top of the panel, and a hole through which marbles were delivered was centered 7 in. from the bottom of the panel. The marbles fell from the hole into a small enclosed box. The box had a clear plastic top and front which enabled *S* to see the marbles but not to handle them.

A mechanism behind the panel provided for the dispensing of the marbles and for the measurement of latency of response. The switch by which *E* turned on the signal light also turned on an electric timer. When *S* pushed one of the knobs, the light was automatically turned off and the timer stopped. In addition, the knob dislodged a marble which fell down a chute into the box.

Procedure.—The *S* was seated in front of the apparatus and was told that he was to play a game. The *E* demonstrated the apparatus and said, "When the light comes on, you push one of the knobs. If you push the correct knob a marble comes out here like this. Now every time the light comes on you push the knob that you think will get you the marble. Remember, just push one knob each time the light comes on." The *S* was told that he was to get as many marbles as he could and when the game was over he could choose two toys from a selection of toys including balloons, plastic figures, etc., which *E* showed him. The *E* took a position behind the apparatus and did not interact further with *S* during the experiment. A second *E* seated in a distant part of the room behind *S* recorded the responses and their latencies.

Three conditions which differed in the percentage of reinforcement of correct response were used. For each *S*, one of the three knobs (either L, M, or R) was designated as the correct knob. The particular knob that was correct yielded reinforcement; choices of the other two knobs were never reinforced. In the three conditions the correct knob yielded 100%, 66%, and 33% reinforcement, respectively. In the 66% and 33% conditions the trials on which a choice of the correct knob was reinforced were determined by a prearranged random schedule. The schedule was utilized in such a fashion that it insured *S* would receive the desired percentage of reinforcement of correct responses. Fifteen *S*s were assigned at random to each of the three conditions. The *S*s were given 80 trials and at the completion of the experiment *S* was allowed to select his prizes.

Results

The three groups differed consistently throughout the 80 trials in the incidence of correct response, as may be seen in Fig. 1. An analysis of variance revealed that the differences among the groups were significant at beyond the .01 level ($F = 7.39$).

The general tendency seen in Fig. 1 is for the frequency of correct response to increase for a short period and then to level off. The 33% group did not show a significant improvement in performance between the first 20 and last 20 trials ($t = 1.66$, $P > .05$). The change in performance from the first 20 to the last 20 trials was significant for both the 66% ($t = 4.00$, $P < .01$) and the 100% group ($t = 2.97$, $P < .01$).[3]

The values of the asymptotic level of response, as estimated from performance on the last 20 trials, were .97, .79, and .53 for the 100%, 66%, and 33% groups, respectively. The values are quite similar to those obtained by Neimark (**14**), who tested adult *S*s under similar conditions with 100% and 66% reinforcement. The asymptotic levels of response obtained

[3] Forty extinction trials were given to these *S*s immediately after the training trials. During the extinction trials no response was reinforced. The difference in the frequency of choice of the stimulus correct during training between the last 10 acquisition trials and the last 10 extinction trials was significant for the 66% group ($t = 2.38$, $P < .05$) and in the 100% group ($t = 2.51$, $P < .05$), but not for the 33% group ($t = 1.46$, $P > .05$).

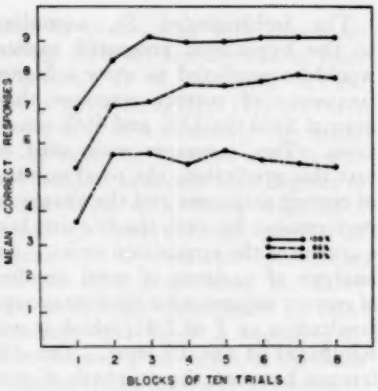

FIG. 1. Mean number of correct responses by normal children for blocks of 10 trials.

by Neimark for the last 20 of 100 training trials were .99 for the 100% group and .83 for the 66% group.

There was a tendency for the average latencies of response in the 100% and 66% groups to decrease during the course of the 80 trials in the manner expected in learning problems of different difficulty. The changes for the 33% group were more erratic and appeared to represent a different process from that of the other two groups. An analysis of variance of the latency scores of the three groups during acquisition is not significant ($F < 1$).

EXPERIMENT II

Goodnow (7), in an analysis of the determinants of choice behavior, has suggested that one of the conditions influencing whether or not *S* will maximize his guesses of the more frequently reinforced stimulus is the level of success *S* will accept in the task. Goodnow suggests that maximizing behavior will be found when *S* will accept less than 100% success as a good final outcome, while other distributions of choices will be found when *S* has an interest in 100% success or in a level of success which is greater than that allowed in the situation.

On the basis of this analysis, it may be hypothesized that different types of behavior will be obtained with *Ss* who differ in the degree of success that they have learned to expect. Normal *Ss*, such as those of Exp. I and of most previous studies, may be assumed to have learned, on the basis of their everyday experience, to expect a high degree of success. Maximizing behavior would not be predicted for these *Ss*. Institutionalized feebleminded children, however, may be assumed to have learned to expect and to settle for lower degrees of success. These *Ss*, therefore, would be predicted to maximize their choices of the reinforced stimulus to a greater degree than would normal *Ss*. It is the purpose of this experiment (*a*) to determine whether institutionalized feebleminded *Ss* do tend to maximize their choices in such a manner, and (*b*) to determine whether their choices of the reinforced stimulus differ significantly from those found for the normal *Ss* in Exp. I.

Method

Subjects.—The *Ss* consisted of 30 feebleminded children chosen at random from among individuals of appropriate MA and CA residing at the Austin State School. The individuals were all of the familial type of mental deficiency and no individuals with gross motor or sensory disturbances were used. The *Ss* were selected so that their average MA would be comparable to that of the normal *Ss* in Exp. I. Previous testing with the Stanford-Binet of more than 80 *Ss* from the population from which *Ss* in Exp. I were selected revealed an average IQ of 119. On the basis of this it was assumed that the *Ss* in Exp. I had an average MA of approximately 6.5 yr. The average MA of the feebleminded *Ss* was 6.1 yr. and their average CA was 12.8 yr.

Apparatus and procedure.—The same apparatus and procedure employed in Exp. I were used. Ten feebleminded *Ss* were tested in each of the three conditions.

Results

The frequencies of correct response made by the three groups of *Ss* are presented in Fig. 2. An analysis of variance of these data indicated that the three groups did not differ significantly in frequency of correct response ($F = 2.91$, $P > .05$). This is in marked contrast with the results for the normal *Ss*, where a highly significant difference among groups was found.

All three groups showed significant increases in frequency of correct response between the first 20 and last 20 trials. The difference is significant at less than the .05 level for the 100% ($t = 2.77$) and 33% groups ($t = 2.48$), and at less than the .01 level for the 66% group ($t = 6.69$).

The normal and feebleminded *Ss* in the 100% groups would not be expected to differ in frequency of correct response during the acquisition trials, for complete success is possible in these groups. The prediction is supported; the averages of 67.5 correct responses for the feebleminded *Ss* and 70.1 for the normal *Ss* do not differ significantly ($t = .33$).

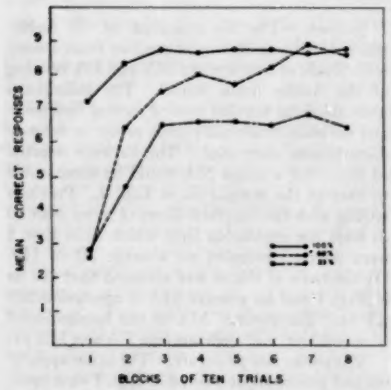

FIG. 2. Mean number of correct responses by feebleminded children for blocks of 10 trials.

The feebleminded *Ss*, according to the hypothesis presented earlier, would be predicted to show a higher frequency of correct response than normal *Ss* in the 33% and 66% conditions. Two measures were used to test this prediction, the total number of correct responses and the change in performance between the first and last quarters of the acquisition series. An analysis of variance of total number of correct responses for the four groups resulted in an F of 2.41, which is not significant at the .05 level. The difference between the numbers of correct responses made during the first and last quarters of acquisition by the four groups of *Ss* were highly significant ($F = 5.35$, $P < .01$). The difference associated with type of *S* was highly significant ($F = 12.74$, $P < .01$). Neither the difference associated with conditions of reinforcement ($F = 3.22$, $P > .05$) nor the interaction between type of *S* and condition of reinforcement was significant ($F < 1.00$). The feebleminded *Ss* had a greater change in performance between the first and last quarters of acquisition than the normal *Ss* in both the 66% ($t = 3.52$, $P < .01$) and, with a one-tailed test, in the 33% group ($t = 1.81$, $P < .05$).

The differences among the average latency scores for the three feebleminded groups are not significant ($F = 1.82$, $P > .05$). An analysis of variance of the latency scores for the normal and feebleminded *Ss* indicated no significant differences among the six groups ($F < 1.00$). The latency curves of the feebleminded *Ss* have the same trends as those of the normal *Ss*.

EXPERIMENT III

The purpose of Exp. III was to determine whether the behavior of normal *Ss* in the task can be influ-

enced by pretraining with games involving different frequencies of reinforcement. It is assumed, in the manner outlined in Exp. II, that there will be less of a tendency for maximizing behavior to occur with *Ss* who have learned to expect high degrees of success than with those who have learned to expect lower degrees of success. It is predicted, therefore, that *Ss* with pretraining on other tasks with high frequencies of reinforcement will show a significantly lower frequency of choice of the reinforcing stimulus than will *Ss* who have pretraining with lower frequencies of reinforcement.

Method

Subjects.—The *Ss* consisted of 30 preschool children attending the same nursery schools as *Ss* in Exp. I. The *Ss* were selected at random from among the children who had not performed in Exp. I. The mean CA was 5.9 yr.

Apparatus.—Four tasks were used—the learning task used in Exp. I and II and three experimental games. The apparatus for the learning task was modified so that a small open box from which *S* could remove the marbles replaced the enclosed box. A marble board, with 50 holes into which *S* could place the marbles, was also used.

The games were the Card Game, the Picture Game, and the Nursery School Game. The *Card Game* apparatus consisted of two sets of cards. The cards were constructed of 2 × 3-in. cardboard on which were mounted small rectangles of black or red paper. One set contained 12 red cards and the other contained 8 black and 4 red cards. The *Picture Game* apparatus consisted of (*a*) 12 cards similar to those in the Card Game, each bearing a rectangle of a different color; (*b*) a panel containing a 7-in. square milk-glass screen; (*c*) a slide projector; and (*d*) 12 slides depicting different animals. The *Nursery School Game* apparatus consisted of (*a*) two green panels 17 in. square, each containing six hooks; and (*b*) 12 pictures of young children mounted in plastic covers with a wire loop at the top of each picture.

Two additional marble boards were also used, one containing 36 holes, and one containing 12 holes.

Procedure.—The *Ss* were presented with the three games followed by the learning problem. The *Ss* were divided at random into two groups of 15 *Ss* each. Group 1 received 100% reinforcement and Group 2 received 33% reinforcement during the games. All *Ss* received 66% reinforcement during the learning problem. The games were dissimilar to each other and to the learning task. Since the games did not involve learning, the different percentages of reinforcement could be given regardless of the specific responses *Ss* made.

The *S* was seated at a small table and was told, "We're going to play several different kinds of games today and you can win some nice prizes." The *E* displayed prizes of the same types used in Exp. I and II, and proceeded with the instructions for the first game. During these instructions *S* was told that he would receive marbles for correct responses and was shown a marble board which he was to fill. The marble boards for the two groups differed; the 36-hole marble board was given to *Ss* in Group 1 and the 12-hole marble board to *Ss* in Group 2. The number of holes in the marble board corresponded to the number of times that *S* would be reinforced during the period in which *S* played the three games. The *S* was further told that when he filled the marble board he would be given another marble board which he could fill and could exchange for his choice of prizes. The use of different marble boards during this period was introduced to insure that the two groups of *Ss*, although receiving different numbers of reinforcements, would have attained the same degree of accomplishment towards their goal. The *S* was then allowed to proceed with the three games. The order in which the three games were played was randomized among *Ss*.

The instructions and procedures for the three games are described below:

In the *Card Game*, one of the decks of cards was placed in front of *S*. He was instructed to draw one card at a time from the top of the deck. He was told that every time he drew a red card he would be given a marble for his marble board. The *S* was allowed to draw 12 cards. The decks were prearranged so that for Group 1 the top 12 cards were red, and for Group 2 the top 12 cards contained 4 red cards and 8 black cards randomly mixed.

In the *Picture Game*, *S* was shown the apparatus and was told that as pictures appeared on the screen he was to select the card that went with the picture. There was no principle by which *S* could select an appropriate card, and reinforcement was given arbitrarily, depending upon the group in which *S* was placed. For Group 1, all of *S*'s responses were reinforced by *E*'s telling *S* that he was correct and giving *S* a marble. For Group 2, 33% of *S*'s responses were reinforced.

The four responses reinforced were determined by a prearranged random sequence.

In the *Nursery School Game*, *S* was told that the panels represented nursery schools and that half of the children in the pictures went to one nursery school and half to the other. The *S* was instructed to place each child in the nursery school which *S* thought the child attended. There was no principle by which the pictures could be separated and again reinforcement was given arbitrarily, depending upon the group in which *S* was placed. All of *S*'s responses in Group 1 were reinforced and 33% of *S*'s responses in Group 2. The responses reinforced in Group 2 again followed a prearranged random sequence.

After completing the games *S* was asked to give the filled marble board to *E* and was given the marble board containing 50 holes. The *S* was told that when this marble board was filled he could select his prizes. The *E* then introduced the learning task. The instructions for this task were identical to those described in Exp. I and II, except that *S* was told to remove the marbles from the box after each reinforced trial and to place them in the marble board. The procedure did not differ for Groups 1 and 2; all *S*s were tested under the 66% reinforcement schedule described in the previous experiments. The *S*s were given 80 trials and were then given their prizes.

Results

The results of the study are presented in Fig. 3. Following the initial 10 trials the groups differed consistently. The *S*s receiving 100% reinforcement on the games preceding the learning problem (Group 1) chose the correct knob less frequently than *S*s who had received 33% reinforcement (Group 2). The correct knob was chosen an average of 44.7 times in Group 1 and 56.6 times in Group 2. The difference is significant at beyond the .025 level with a one-tailed test ($t = 2.11$).

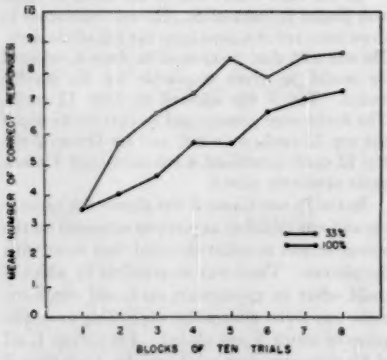

FIG. 3. Mean number of correct responses in blocks of 10 trials.

DISCUSSION

In this study it was assumed that children learn to expect certain frequencies of reinforcement on the basis of their everyday experience, and that children living in a normal, responsive environment develop a higher expectancy of reinforcement than do children living in an institution. It was also assumed that the expectancies that children develop may be modified by manipulating the frequency of reinforcement that the child receives in an experimental setting. It was hypothesized that different types of behavior would be obtained with *S*s who differ in the degree of success that they have learned to expect, and it was predicted that *S*s with low expectancies would show a greater frequency of choice of the reinforcing stimulus than would *S*s with higher expectancies. This would result from the attempt by *S*s with higher expectancies to seek, through variable behavior, a means by which they could obtain a frequency of reinforcement corresponding to the frequency which they expect.

The total number of correct responses made by three groups of normal children receiving different percentages of reinforcement differed significantly, but the number of correct responses made by three groups of feebleminded children tested under the same procedure as the normal *S*s did not. The two types of *S*s differed significantly in rate of learning under both 66% and 33% reinforcement, but did not differ significantly under 100% reinforcement. In learning a response which yielded 66% reinforcement, normal *S*s receiving 100% reinforcement

on pretraining games showed significantly poorer learning than did *Ss* receiving 33% reinforcement on the pretraining games.

In Exp. I it was found that the mean response levels of normal children were similar to those for groups of adult *Ss* given 66% and 100% reinforcement in a similar situation (**14**). The results are also similar to those of Messick and Solley (**13**), who found that children tend to approach an asymptotic level of response in a probability learning situation.

In the 33% and 66% groups of Exp. I, a tendency was found for *Ss* to follow one of two modes of response; they either adopted the correct response relatively quickly or persisted in variable behavior throughout all of the trials. Individual differences in responses on a two-choice task have also been reported by Anderson and Grant (**1**). It is of interest that even *Ss* who persisted in responding in a variable fashion were able to tell *E* at the end of the experiment which knob had yielded reinforcement. The fact that *Ss* knew the correct response but did not make it supports the view that *Ss* with a high expectancy are unwilling to accept a solution which yields low frequencies of reinforcement.

There are several alternative interpretations of the differences in behavior of the normal and feebleminded children which differ from the one previously presented. One that seems quite reasonable is that institutionalized children may have less experience in playing games than normal children, hence are less likely to know the nature of such games. If this were the case, their behavior might be similar to that of adult *Ss*, who, when tested in an obscure task, tend to depart from the probability matching behavior found in less obscure tasks (**19**). A second alternative interpretation is that feebleminded children may tend to choose the reinforced stimulus more frequently than do normal *Ss* because they are more rigid and tend to perseverate more than do normal *Ss*. This interpretation does not, however, appear to be tenable in light of the results of several recent studies in discrimination learning which show that normal and feebleminded *Ss* of the same MA do not differ in rigidity of response or in learning speed (**16, 17**).

The results of Exp. III also provide support for the hypothesis that behavior in probability learning tasks is dependent upon *Ss'* expectations concerning level of reinforcement. The fact that pretraining with different frequencies of reinforcement affects behavior in a probability learning task is in accord with the results of experiments with adult *Ss* where the probability of reinforcement is shifted during the course of training. In these studies (**5, 8, 15**), differences in behavior have been found as a function of the shift in probabilities of reinforcement.

The form of the curves for the two groups in Exp. III tends to differ from that of the group in Exp. I which also received 66% reinforcement. The differences in the curves may be related to the differences in procedures used in the two experiments. In Exp. III, *Ss* were given a more defined goal than were *Ss* in Exp. I. The use of a more defined goal may lead to a greater tendency for *Ss* to maximize their choices of the reinforcing stimulus.

Summary

Three experiments are reported which investigate the performance of children in a probability learning task. In Exp. I and II, normal and feebleminded children were tested on a problem where the correct response resulted in either 100%, 66%, or 33% reinforcement. In Exp. III normal children were given pretraining on three nonlearning games with 100% or 33% reinforcement, and were then trained on the discrimination problem with 66% reinforcement of the correct response. Predictions were made from an hypothesis relating performance to *Ss'* expected frequency of reinforcement. Feebleminded *Ss* performed at a higher level in the 33% and 66% conditions than normal *Ss* and at a comparable level in the 100% condition. The *Ss* in Exp. III receiving 100% reinforcement during pretraining made significantly fewer correct responses in the learning problem than did *Ss* receiving 33% reinforcement during pretraining.

References

1. Anderson, N. H., & Grant, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.
2. Brand, H., Sakoda, J. M., & Woods, P. J. Contingent partial reinforcement and the anticipation of correct alternatives. *J. exp. Psychol.*, 1957, **53**, 417–424.
3. Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–104.
4. Estes, W. K. Individual behavior in uncertain situations: An interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954.
5. Estes, W. K., & Straughn, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
6. Gardner, R. A. Probability learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.
7. Goodnow, J. J. Determinants of choice distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.
8. Goodnow, J. J., & Pettigrew, T. F. Effect of prior patterns of experience upon strategies and learning situations. *J. exp. Psychol.*, 1955, **49**, 381–389.
9. Goodnow, J. J., & Postman, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.
10. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
11. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
12. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternatives. *J. exp. Psychol.*, 1951, **41**, 291–297.
13. Messick, S. J., & Solley, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.
14. Neimark, E. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.
15. Parducci, A. Alternative measures for the discrimination of shift in reinforcement-ratio. *Amer. J. Psychol.*, 1957, **70**, 194–202.
16. Plenderlith, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *J. genet. Psychol.*, 1956, **88**, 107–112.
17. Stevenson, H. W. & Zigler, E. F. Discrimination learning and rigidity in normal and feebleminded individuals. *J. Pers.*, 1957, **25**, 699–711.
18. Straughn, J. H. Human escape learning in relation to reinforcement variables and intertrial conditions. *J. exp. Psychol.*, 1956, **52**, 1–8.
19. Wyckoff, L. B., & Sidowski, J. B. Probability discrimination in a motor task. *J. exp. Psychol.*, 1955, **50**, 225–231.

(Received September 13, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE INTERACTION OF ABILITY AND AMOUNT OF PRACTICE WITH STIMULUS AND RESPONSE MEANINGFULNESS (m, m') IN PAIRED-ASSOCIATE LEARNING [1]

VICTOR J. CIEUTAT, FREDRIC E. STOCKWELL

Louisiana State University

AND

CLYDE E. NOBLE

Montana State University

The present program on the role of meaningfulness (m) in verbal learning began with the development of the m scale of association frequency (**7**), and proceeded to experimental studies of the effects of this variable upon serial learning (**8**) and paired-associate learning (**10**). Using either method, *Ss*' rate of memorizing two-syllable nouns and paralogs is a positive function of m.[2] The latter two investigations also discovered interactions between *Ss*' initial ability and their reactivity to differences in m. In all previous studies, however, the m values of the stimuli and responses were deliberately covaried, so the relative contributions of the S and R terms to the variance in performance have not yet been measured. There exists the possibility of an S × R interaction with respect to m, as well as interactions with amount of practice if wide ranges in proficiency are produced.

Variants of the m scale have also been applied to consonant-vowel-consonant or "CVC" material (**11**) with similar effects on paired-associate acquisition (**4, 5, 6, 11, 13, 14**). In this second group of experiments the association values of S and R were separately manipulated in two cases (**5, 13**), but no attempt was made simultaneously to evaluate the role of individual differences. To determine the extent to which initial ability interacts with the locus of meaningfulness in paired-associate CVC learning, the S and R terms must be independently varied with respect to this factor. Because of its methodological status, known reliability, and current relevance to the population available to the writers, the m' scale (**11**) will be employed for the second objective. This x/σ scale consists of a random sample of 100 CVC syllables, rated for number of associations by 200 *Ss*, whose measurements were transformed into deviates of the normal curve by the method of successive intervals. It has been shown (**11**) that the relationship between m and m' is positive, thus rationalizing

[1] This paper is based upon research performed by the junior authors under the guidance of the senior author. Experiment I was a 1957 undergraduate project by Stockwell reported to the American Psychological Association, Sept. 1957. Experiment II was a 1957 master's thesis by Cieutat reported to the Southeastern Psychological Association, April 1958. We are grateful to Linda Cieutat and D. R. Gannon for checking the statistical analyses, and to B. I. Murstein and B. J. Underwood for helpful suggestions.

[2] These findings have been confirmed for serial learning by Dowling and Braun (**1**), and for paired-associate learning by Kimble and Dufort (**2**). As a further index of the m scale's reliability (**10**, p. 16), Rocklyn, Hessert, and Braun (**12**) have shown that the defining operations (**7**) yield measurements which are practically invariant over age differences from 20 to 66 yr.: $r = .96$.

the definition of m' as an estimate of m. This implies that both operations reflect quantitative changes in the same hypothetical concept: the number of habit connections evoked by a given stimulus (7).

The two experiments described here will investigate (*a*) the nature of S and R combination in paired-associate learning under manipulations of m and m', (*b*) the interaction of each of these measures with amount of practice, and (*c*) the role of initial ability in determining the type and extent of such interrelationships when m' is varied. Experiment I uses the m scale and the pronunciation method, whereas Exp. II uses the m' scale and the spelling method.

Experiment I

After considering evidence from earlier studies of the influence of meaningfulness, Noble and McNeely proposed the S-R motor-patterning hypothesis that "differential facilitation is primarily connected with responding (articulation) and that stimulus factors are of secondary importance" (10, p. 21). The following predictions were made for a 2 × 2 factorial design employing four independent groups of *Ss* practicing homogeneous paired-associate lists: "(*a*) the simple effect favoring high S-high R meaningfulness (e.g., m = 6–9) over low S-low R meaningfulness (e.g., m = 0–2) will be greater than for any other comparison, and (*b*) the difference in the main effect of m on the response side will be greater than the comparable difference on the stimulus side. Given (*a*) and (*b*), the likelihood of intrinsic S-R interaction with respect to m is slight" (10, p. 21).

The last sentence reveals the uncertainty of current knowledge about the nature of S and R combination in paired-associate learning. If the relationship is intrinsically *non-additive* when m is varied, and if this interaction should itself interact with number of trials (N), then it is conceivable that the S × R variance in a simple factorial experiment might be averaged out in criterion scores due to ignoring the trend factor N. Data reported in an unpublished study by Sheffield (13) indicate such a possibility, as pointed out earlier (8, p. 443).

For this reason, the present experiment will test these implicates in an S × R × N mixed factorial design (3, Type III), thereby jointly evaluating the three-factor interaction in addition to the main effects and two-factor interactions. As far as the basic issue underlying so-called "perceptual" (2, p. 363) and "motor" (10, p. 22) principles of verbal learning is concerned, it is evident that when the N classification is disregarded this more complex design includes the 2 × 2 factorial conditions proposed above.

Method

Apparatus.—The learning device consisted of a Patterson memory drum set to present the paired associates at a 4-sec. rate with a 4-sec. intertrial interval.[3] The S term was exposed for 2 sec., followed by S and R together for 2 sec. An idler attachment permitted four sets (trials) of 10 pairs to be presented in series to *S* without changing tapes. Six tapes were constructed, each containing a practice and an experimental list typed in capital letters. There were four different random trial sequences for each list, and the series was repeated on Trials 5 and 9. On each tape the practice list appeared on the right side, the experimental list on the left. Sliding shutters over the aperture of the memory drum permitted one list to be covered while the other was being presented.

For the *practice* list, 10 pairs of three-syllable adjectives were used as in a previous study (10), the four trial sequences being the same for all *Ss*. This practice period was intended to reduce variability due to learning-to-learn and to provide comparisons of the initial ability and predictability of the four groups. For the *experimental* lists, 40 items were drawn from the high (H) and low (L) extremes of the m scale (7) such that the mean median m value of the 20 H items was 7.50 and that of the 20 L items was .29. Each of these lists consisted of 10 pairs of two-syllable nouns or paralogs, combined so as to avoid mediated associations and similar sources of bias (10).

Subjects.—The *Ss* consisted of 80 undergraduate students at Louisiana State University. There were 44 women and 36 men, whose ages ranged from 18 to 31 (mean = 20.7 yr.). All were naive with respect to paired-associate learning.

Procedure.—Four experimental groups were formed in terms of the possible combinations of the m values of the S and R items: H-H, L-H, H-L, and L-L. In Groups L-L and H-H the pairs were reversed for half the *Ss*, and the pairs learned by Group L-H were reversed for Group

[3] Photographs of this device appear in Wickens and Meyer (15, p. 59).

H-L, giving six lists in all. There were 20 *S*s in each group, assigned consecutively by fives to the six lists in a counterbalanced order. Each time a list was repeated, it began with a different trial sequence. Thus, all four sequences appeared equally often for the four experimental groups.

Each *S* received 12 trials on the practice list. After a 2-min. rest, this was followed by 12 trials on the appropriate experimental list. During the rest period between lists *E* engaged *S* in conversation. The experiment was performed in an air-conditioned laboratory, so the temperature and humidity conditions were quite constant.

The instructions to *S*s were the same as in the earlier study (10). On the first presentation of the list (Trial 1), *S* merely read the paired words aloud and standard pronunciations were established. Attempted anticipations of the correct responses began on the second presentation (Trial 2). The correction procedure was used, and both S and R were pronounced.

Results

To evaluate the comparability of the four groups before introduction of the experimental treatments, a simple 4×1 analysis of variance was performed upon the total number of correct responses (R+) made during Trials 2–12 on the practice list. The *F* ratio is only .13, which for $df = 3/76$ is not significant. The experimental groups, therefore, may be considered to be of equal initial ability.

Table 1 presents a 2×2 analysis of the average performance of each group

TABLE 1

MEAN TOTAL CORRECT RESPONSES (R+) DURING TRIALS 2–12 IN EXP. I

m Value of S Term	*m* Value of R Term		Row Means	S-Difference
	L(0.29)	H(7.50)		
L(0.29)	34.00	85.40	59.70	
H(7.50)	71.60	92.75	82.18	22.48
Column Means	52.80	89.08		
R-Difference	36.28			

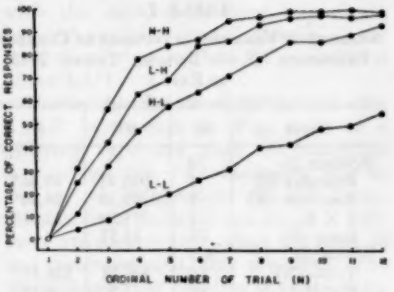

FIG. 1. Acquisition curves for lists of 10 paired associates as a function of practice (N) in Exp. I. The four S-R combinations of low (L) and high (H) meaningfulness (*m*) represent the parameter. Each curve contains 20 *S*s.

on the experimental lists. The order of increasing difficulty in terms of cell means is: H-H, L-H, H-L, L-L. From the main effects of the S and R terms, it can be seen that the average difference due to R is greater than that due to S.

Figure 1 shows the percentage of correct responses as a function of practice, with type of S-R pair as the parameter. The curves are quite regular with no inversions during the entire training period. There is a tendency for negative acceleration to occur when the R term has high *m* value, and for the curves to be linear or slightly positively accelerated when the R term has low *m* value. Group L-L is markedly inferior to the other three groups, but it seems to be approaching the same asymptote.

A $2 \times 2 \times 11$ analysis of variance of the number of correct responses made by the four groups during Trials 2–12 is summarized in Table 2. All main effects and interactions are significant. In variance terms, the relative magnitudes of the *inter*-individual comparisons are: $\sigma^2_R > \sigma^2_S > \sigma^2_{SR}$. For the *intra*-individual comparisons the order is: $\sigma^2_N > \sigma^2_{SRN} > \sigma^2_{RN} > \sigma^2_{SN}$. It is interesting to

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RESPONSES (R+) DURING TRIALS 2–12 IN EXP. I

Source	*df*	*MS*	*F*
Between *Ss*	79		
Stimulus (S)	1	918.41	49.46*
Response (R)	1	2392.50	128.84*
S × R	1	415.94	22.40*
Error (*b*)	76	18.57	
Within *Ss*	800		
Trials (N)	10	400.61	328.37*
S × N	10	5.10	4.18*
R × N	10	8.42	6.90*
S × R × N	10	8.55	7.01*
Error (*w*)	760	1.22	
Total	879		

* $P < .001$.

observe that the variance of R exceeds the variance of S, and that R interacts more with N than S does.

To determine the predictability of each group from initial ability measures, product-moment correlations (r) were computed between the total correct responses (R+) on the practice list and the same score on the experimental lists. The correlations are as follows: .45 (H-H), .51 (L-H), .72 (H-L), and .81 (L-L). This systematic progression could mean that validity increases with difficulty, but the observed trend is probably due merely to greater restriction of range in the easier lists. The average r calculated by Fisher's Z transformation is .65, which is consistent with the .50 reported previously (10) using trials-to-criterion scores.

Discussion

Sheffield's (13) experiment is rather closely akin to Exp. I. He employed so-called "nonsense" CVC's and familiar three-letter words to constitute two levels of high (H) and low (L) "meaningfulness." The *Ss* practiced heterogeneous lists containing the four possible S-R combinations of the H and L items: H-H, L-H, H-L, and L-L. Both S and R were pronounced and 10 trials were given. The pairs were learned with the increasing rank order of difficulty stated above. Regardless of the nature of the S term, greater over-all proficiency was attained when the R term was an actual word than when it was a CVC syllable. Specifically, the difference between the mean levels of the two high-R and two low-R conditions exceeded the difference between the mean levels of the respective high-S and low-S condition. This implied the dominance of response factors over stimulus factors in paired-associate learning.

Unfortunately, Sheffield did not report a statistical test of S × R interaction, due probably to nonindependence of the four conditions. Graphs of his *Ss*' curves indicated, however, that the *initial* separation of H-H and L-H was greater than that of H-L and L-L but tended to become equalized with practice. This suggests the possibility of interaction, but artifacts could have resulted since the H and L items were not actually quantified by means of common operations. Reviewing this experiment in 1952, Noble (8) proposed using the *m* scale in an appropriate design to determine whether a significant S × R interaction is present.

Turning now to the present data, we recall that the first prediction from the S-R motor-patterning hypothesis (10) was that there would be a greater difference between Groups H-H and L-L than between any other two groups. This is confirmed by Table 1 and Fig. 1, and is in general agreement with five other paired-associate experiments (6, 10, 11, 13, 14). Two exceptions are a study by Kimble and Dufort (2, Exp. II), in which the order of proficiency was L-H, H-H, H-L (no L-L condition being given), and a study by Mandler and Campbell (5, Exp. I), in which the order was H-H, L-L, L-H, H-L. Both sets of results, however, are difficult to compare with Table 1. Aside from the complications of a confounded design (10, p. 21), half the Kimble-Dufort items were unscaled three-letter words, and the Mandler-

Campbell CVC data came from three different groups of *Ss* ($N = 12$) unmatched for ability. In a more adequate follow-up study (**5**, Exp. II) using a larger sample ($N = 30$) and a constant-trials design, however, Mandler and Campbell obtained the following ranking: H-H, H-L, L-H, L-L. The extremes are as predicted but the middle groups are reversed.

The second expectation (**10**) was that the variance attributable to R would be greater than that attributable to S. This is confirmed by Table 1, where the main effect (mean difference) of R is 36.28 correct responses vs. 22.48 for the main effect of S. Although Sheffield's (**13**) findings agree, the Mandler-Campbell (**5**) data do not; and the Kimble-Dufort (**2**) design is inappropriate. Such equivocal reports are not surprising in view of the methodological differences among these experiments. This constitutes one reason for performing Exp. II below. From Table 2 it may be seen that the size of σ^2_R is approximately 2.6 times that of σ^2_S, indicating that the meaning-learning correlation is primarily a response effect and that the stimulus factor is secondary. If our rationale linking m and m' is correct (**7, 11**), a similar phenomenon ought to occur in the learning of CVC pairs.

The remaining issue concerns the interactions. According to Fig. 1 and Table 2, there is a significant tendency for an increase in the m value of the S term to cause a greater difference in rate of acquisition when the R term has low m value than when the R term has high m value. Referring to Table 1, the difference between the simple effects (cell means) of Groups L-L and H-L is 37.60 correct responses whereas the difference between Groups L-H and H-H is only 7.35. (There is a related tendency for increases in R to produce more facilitation when S is low than when it is high: 51.40 vs. 21.15 correct responses.) In Fig. 1 these interacting trends appear as differential acquisition rates, and Table 2 shows that both S and R as well as S × R interact significantly with amount of practice (N). Consistent with the motor-patterning hypothesis, but not explicitly stated therein, is the fact that $\sigma^2_{RN} > \sigma^2_{SN}$. The ratio is about 1.6/1.

Unlike Sheffield's (**13**) curves, the S × R interaction in Fig. 1 is of a different type and does not disappear late in practice. Furthermore, due to the use of independent *Ss* and homogeneous lists, the level and range of difficulty are greater than when the same *Ss* and heterogeneous lists are used (**10**). This leads to the problem of *extrinsic* vs. *intrinsic* interaction (**3**, p. 124). In view of the fact that all our groups were well matched for ability, the *Ss* tested individually, procedural factors equalized for each group, and the H and L items quantified by means of common operations, it is unlikely that the significant S × R interaction is caused by extraneous decremental factors associated, say, with Group L-L and not with the other groups. We conclude, therefore, that the observed interaction is intrinsic; i.e., due only to the treatments and not to uncontrolled factors.

Experiment II

Our analysis of experiments using CVC syllables indicated that inadequate attention had been paid to the importance of initial ability (A) in determining the influence of S and R meaningfulness. There is also a need for greater procedural comparability between "verbal" and "syllable" experiments than has existed up to now.

In order to perform a more comprehensive test of the Noble-McNeely hypothesis (**10**), Exp. II will expand the 2 × 2 × 11, S × R × N design of Exp. I to a 3 × 3 × 20 × 2, S × R × N × A mixed factorial (**3**, Chap. 13), and apply it in the context of paired-associate CVC learning. A prediction favoring the facilitative effect of the R term over that of the S term, when high (H), medium (M), and low (L) levels of m' are used, suggests the following rank order of proficiency for the nine experimental groups disregarding ability: H-H, M-H, L-H, H-M, M-M, L-M, H-L, M-L, L-L. Assuming next that the same hypothetical factors are operating in learning items from the m' scale as from the m scale, *Ss'* level of pre-experimental ability (A) should interact with the S, R, and N factors (**8, 10**), and the main effects and interactions observed in

Exp. I will be given an opportunity to be confirmed or refuted.

Method

Apparatus.—The memory drum and time constants were the same as in Exp. I. A *practice* list was constructed of 10 pairs of CVC syllables selected from Mandler's (4) table to avoid overlapping with the m' scale. The *experimental* lists were taken from the Noble-Stockwell-Pryer m' norms (11). Both the practice list and the experimental lists were presented in five random trial sequences and the series was repeated on Trial 6. The practice period was intended to reduce variability due to learning-to-learn and to serve as an estimate of initial ability as in Noble and McNeely (10) and in Exp. I.

Nine tapes were prepared, each having the 10-pair practice list and one of the 10-pair experimental lists typed in capital letters. The mean m' values of the S and R terms in the experimental lists varied from low (L = $\overline{m}'$ of .68) to medium (M = $\overline{m}'$ of 1.72) to high (H = $\overline{m}'$ of 2.91) in all possible combinations: H-H, M-H, L-H, H-M, M-M, L-M, H-L, M-L, and L-L. Lists H-M, H-L, and M-L were composed of pairs formed by reversals of the items in the M-H, L-H, and L-M lists, respectively.

The following criteria governed the construction of the lists: no consonant or vowel letters were duplicated in any pair (e.g., *MOF-DOQ*); no alphabetical sequence of consonants occurred within a single syllable (e.g., *MEN*) or from syllable to syllable in any pair (e.g., KEP-QAD); and no pairs constituted meaningful associations (e.g., BOR-DAL, GUD-YIR). In the randomization of the lists the following criteria were adopted: each list was considered a circular series and no pair appeared twice within any five consecutive positions (i.e., there were at least four intervening pairs before a repeat); no pair occupied the same serial position on any two adjacent lists; and no two pairs were in the same relative positions on two adjacent lists.

Subjects.—The *Ss* consisted of 90 undergraduate students at Louisiana State University. There were 35 men and 55 women, whose ages ranged from 17 to 27 (mean = 19.9 yr.). All were naive with respect to paired-associate learning. On the basis of the total number of correct responses (R+) made on the practice list, *Ss* were classified as either Fast (R+ $\geq$ 6) or Slow (R+ $\leq$ 5) learners (N = 45 each). A pilot study was conducted to determine the median (cut-off) score on this list. The mean of the Slow group was 2.0 R+'s, while that of the Fast group was 12.5 R+'s.

Procedure.—Each *S* received 10 trials on the practice list. After a 2-min. rest, this was followed by 20 trials on the appropriate experimental list. During the rest period between lists, *E* engaged *S* in conversation. Five *Ss* in each of the two ability levels (A) were then randomly assigned to each of the nine lists in a counterbalanced order of trial sequences. To effect the matching, *E* had to change tapes occasionally and to reject a number of *Ss*.

The instructions to *Ss* were similar to those in prior studies (10, 11), especially Exp. I. The correction procedure was used, and both S and R were spelled aloud.

Results

To evaluate the comparability of the nine groups within each A level before introduction of the experimental treatments, simple 9 × 1 analyses of variance were performed upon the total number of R+'s made during Trials 1–10 on the practice list. Both F ratios are less than unity (df = 8/36). Even the combined F ratio is only .34, which for 8/81 df is not significant. The experimental groups, therefore, may be considered to be of equal initial ability, either preserving or disregarding the Slow vs. Fast classification. Differences in ability between men and women were also tested by comparing their cumulated R+ scores on the practice list. The t ratio was 1.16 for 88 df, which is not significant. The average product-moment r between total R+'s on the practice list and the same measure on each of the experimental lists is .63. This correlation is significant at the 5% point for 8 df. For computation of the individual r's the scores of both ability groups were combined (N = 10) for each experimental list. Fisher Z transformations were then used to compute the average of the nine r's. The predictability of the present data is very similar to that reported previously (10) and in Exp. I.

Table 3 presents a 3 × 3 analysis of the average performance of each group on the experimental lists. From the main effects of the S and R terms,

TABLE 3

MEAN TOTAL CORRECT RESPONSES (R+) DURING TRIALS 1–20 IN EXP. II

m′ Value of S Term	m′ Value of R Term			Row Means	S-Difference (H-L)
	L(0.68)	M(1.72)	H(2.91)		
L(0.68)	6.0	16.7	30.5	17.7	
M(1.72)	17.4	31.4	77.5	42.1	23.3
H(2.91)	21.0	47.2	54.8	41.0	
Column Means	14.8	31.8	54.3		
R-Difference (H-L)	39.5				

exactly as in Table 1, it is clear that the average difference due to R is greater than that due to S.

A $3 \times 3 \times 4 \times 2$ mixed factorial analysis of variance (**3**) was computed on the R+ scores of each subgroup ($N = 5$), arranged in four blocks of five trials each. The four factors in this $S \times R \times N \times A$ design were stimulus meaningfulness (m') at three levels (H, M, L), response meaningfulness (m') at the same three levels (H, M, L), number of practice blocks (N) at four levels, and initial ability (A) at two levels (Slow, Fast). There were thus $9 \times 2 = 18$ independent treatment groups of 5 *Ss* each practicing for 20 trials. The summary analysis is shown in Table 4.

It is evident that all main effects (S, R, N, A), all two-factor interactions ($S \times R$, $S \times A$, $R \times A$, $S \times N$, $R \times N$, $N \times A$), and all three-factor interactions except $S \times R \times A$ ($S \times R \times N$, $S \times N \times A$, $R \times N \times A$) are significant; but the four-factor $S \times R \times N \times A$ interaction is not. These results confirm all of the findings in Exp. I which pertain to the m scale and extend them to the domain of the m' scale. In addition to verifying the main effect of A, the three two-factor interactions involving A and the three three-factor interactions involving N clearly reflect the intimate relationships existing between initial ability and amount of practice in paired-associate learning when meaningfulness is independently varied. In variance terms, the relative magnitudes of the significant *inter*-individual comparisons are: $\sigma^2_A > \sigma^2_R > \sigma^2_S > \sigma^2_{RA} > \sigma^2_{SA} > \sigma^2_{SR}$. These inequalities agree perfectly with Exp. I. For the significant *intra*-individual comparisons the order is: $\sigma^2_N > \sigma^2_{NA} > \sigma^2_{RN} > \sigma^2_{SN} > \sigma^2_{RNA} > \sigma^2_{SNA} > \sigma^2_{SRN}$. These latter variances agree with Exp. I except for the placement of σ^2_{SRN}, a point to which we shall return later.

The major hypothesis generalized from the previous study (**10**) was tested by computing a rank-difference correlation coefficient between the predicted order of proficiency and the

TABLE 4

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RESPONSES (R+) DURING FOUR 5-TRIAL BLOCKS IN EXP. II

Source	*df*	*MS*	*F*
Between *Ss*	89		
Stimulus (S)	2	1424.32	16.30***
Response (R)	2	2934.51	33.59***
Ability (A)	1	4396.01	50.31***
S × R	4	348.87	3.99**
S × A	2	417.22	4.78*
R × A	2	1025.14	11.73***
S × R × A	4	132.36	1.51
Error (*b*)	72	87.37	
Within *Ss*	270		
Trial blocks (N)	3	2340.84	208.26***
S × N	6	116.18	10.34***
R × N	6	151.63	13.49***
N × A	3	246.43	21.92***
S × R × N	12	25.90	2.30**
S × N × A	6	37.88	3.37**
R × N × A	6	48.69	4.33***
S×R×N×A	12	11.47	1.02
Error (*w*)	216	11.24	
Total	359		

* $P < .025$.
** $P < .01$.
*** $P < .001$.

obtained order. The latter was based on the cell means of Table 3 ($N = 10$), disregarding the ability classification. For 8 *df*, rho = .88, which is significantly greater than zero at the 1% point. Taking account of individual differences, the motor-patterning hypothesis is about equally tenable for high and low A levels. The two rho values ($N = 5$) are .70 for the Slow group and .75 for the Fast group, both of which are significant at the 5% point.

Discussion

The present results generally confirm earlier studies in this series with respect to the meaning-learning correlation (**8, 10, 11**), the differential importance of S and R and their nonadditive combination (Exp. I), the interaction of meaningfulness with initial ability (**8, 10**), and the dependence of these various relationships upon amount of practice (**8, 10, 11**). With respect to the four combinations of H and L tested in Exp. I, the rank-order agreement in Tables 1 and 3 is perfect. This fits with the Sheffield (**13**) data, but conflicts with the Kimble-Dufort (**2**) and Mandler-Campbell (**5**) studies mentioned above. In the light of previous methodological comments, however, one should not take these disagreements too seriously. For instance, the Mandler-Campbell experiments (**5**, Exp. I vs. Exp. II) do not even agree with each other. Using rank-correlation again, the cell means for the two sets of nine combinations of H, M, and L in their paper yield a rho of only .03. It should be added, however, that both sets of data are positively, though not significantly, related to our hypothesized rank order of the nine groups: rho $\approx$ .60, which for 8 *df* fails to reach significance at the 5% point.

The correspondence between fact and hypothesis in Table 4 (rho = .88) is not as close as expected by the writers. One curious inversion is the superiority of the M-H group over the H-H group. Upon investigation of this discrepancy, it was found to be an experimental artifact caused by a slightly greater average m' value for the R term in the M-H list ($\bar{m}' = 3.04$) than in the H-H list ($\bar{m}' = 2.79$). The mean difference in m' scale values of .25 is not significant ($t = 1.51$; $df = 18$; $p > .10$), but apparently it was powerful enough to produce the obtained reversal. There is no a priori reason to expect any intrinsic superiority for the M-H combination.

Aside from testing the S-R motor-patterning hypothesis at more points along the meaningfulness continuum, the present study was designed to clarify the role of individual differences. The predicted interaction of the A factor was found with S as well as with R, and the magnitude of σ^2_{RA} exceeded that of σ^2_{SA}. Because Exp. I found the ratio σ^2_R/σ^2_S to be approximately 2.6/1 and the ratio $\sigma^2_{RN}/\sigma^2_{SN}$ to be 1.6/1, it is interesting to note that in Table 4 the analogous σ^2_R/σ^2_S ratio is about 2.1/1 and the quantity $\sigma^2_{RN}/\sigma^2_{SN}$ is roughly 1.3/1. The inference is clear that at least twice the variance in performance due to meaningfulness is attributable to R as to S, and that nearly one and a half times the interaction variance with practice can be similarly identified. That S $\times$ R interacts more with N in Exp. I than in Exp. II is probably due to the greater ease of learning m scale items; i.e., the general law covering all types of material is S $\times$ R $\times$ N.

In previous experiments (**8, 10**) the significant interactions between m and A suggested that "slow learners are more sensitive to differences in meaningfulness than are relatively fast learners" (**10**, p. 19). The present situation, however, produces trends which are exactly the opposite. If one calculates the mean number of R+'s during Trials 1–20 for the Slow and Fast *S*s in the L-L, M-M, and H-H groups, it appears that Fast *S*s profit more by increases in m' than do Slow *S*s. The average correct anticipations increase with m' as follows: Slow (5.0, 24.8, 20.2) vs. Fast (7.0, 38.0, 89.4). Rather than assigning this reversal to scale differences between m and m', or perhaps to other procedural factors such

as pronouncing vs. spelling, the writers believe that the principal variable is the extreme difficulty of the CVC syllable learning task, as mentioned earlier (11). Comparing the present H-H, L-H, H-L, and L-L conditions with those in Exp. I, one finds that the acquisition curves during Trials 1–12 proceed more slowly in every case in this experiment. A similar comparison was true of two earlier studies (10, 11) which did not involve independent variation of S and R, and in which all pairs were learned by the same *S*s.

The foregoing discussion of individual differences may be summarized by the principle that relatively difficult paired-associate material (e.g., CVC syllables) produces *diverging* slopes for Fast vs. Slow *S*s, whereas easier material (e.g., dissyllabic words) produces *converging* slopes. The general curve form is of the sigmoid family. If one now imagines a graphical plot of $R\% = f(m')$ with A as the parameter, it is clear why the form of the meaning-ability interaction should differ in the two experimental situations. A comprehensive investigation of this problem, employing several levels of A and either a wider range of m' or greatly extended practice, is needed. Since the effects of familiarization on *serial* learning (9) imply that variations in the m (hence m') value of paired associates must also reflect the transfer of prior frequencies of experience (n), the suggested experiment might reveal that meaningfulness is a secondary phenomenon derivable from the same basic variable which controls all learning. This coordination would be demonstrated neatly if one could produce similar acquisition curves by manipulating either n or m'. The empirical question is whether familiarization would affect the S and R terms differentially. In 1921 Winzen (16) reported the S-effect to be greater than the R-effect, but Sheffield (13) got the reverse; Mandler and Campbell (5) found no differences. Greater amounts of familiarization in the latter study, however, might have produced significant facilitation. Available data (9) indicate a continuous function.

Summary

In Exp. I, four independent groups of 20 *S*s practiced homogeneous lists of 10 pairs of dissyllables for 12 trials. Two levels of meaningfulness (m) were employed, high ($H = 7.50$) and low ($L = .29$), giving the following stimulus-response m values: H-H, L-H, H-L, and L-L. In agreement with prediction, the lists were learned with the rank order of proficiency shown above when practice (N) was varied. The mean total correct responses made by each group during Trials 2–12 were 92.75, 85.40, 71.60, and 34.00, respectively. Greater facilitation was attributable to the m value of the R term than to the m value of the S term, and there were significant S × R, R × N, S × N, and S × R × N interactions.

Experiment II evaluated the influence of scaled meaningfulness (m') upon paired-associate syllable learning together with the interaction of the S and R factors with initial level of ability (A) and amount of practice (N). Ninety *S*s first received 10 trials on a practice list used to classify them as Fast or Slow on the A factor. This was followed by 20 trials on one of nine experimental lists in which the m' values of S and R were independently varied at three levels: .68, 1.72, and 2.91.

A rho of .88 was obtained between the observed rank order of experimental groups and that expected from the hypothesis favoring the influence of the R term over that of the S term. The effects of S, R, N, and A were significant, as were all two-factor interactions and all three-factor interactions except S × R × A. The relative magnitudes of the variances in Exp. I and Exp. II agreed in six out of seven cases.

From these two experiments it was concluded that at least twice the variance in performance due to meaningfulness (m or m') is attributable to R as to S, and that S-R combination in paired-associate learning is intrinsically nonadditive. The degree to which learning ability interacts with meaningfulness (m or m') depends upon the difficulty of the material and the amount of practice. The results generally confirm earlier experiments in this series, thus supporting the S-R motor-patterning hypothesis. Contradictory findings of recent studies were discussed.

References

1. Dowling, R. M., & Braun, H. W. Retention and meaningfulness of material. *J. exp. Psychol.*, 1957, **54**, 213–217.
2. Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.

3. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
4. MANDLER, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. *Amer. J. Psychol.*, 1955, **68**, 662–665.
5. MANDLER, G., & CAMPBELL, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.
6. MANDLER, G., & HUTTENLOCHER, J. The relationship between associative frequency, associative ability and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.
7. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
8. NOBLE, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; Erratum 1952, **44**, 465.
9. NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.
10. NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.
11. NOBLE, C. E., STOCKWELL, F. E., & PRYER, M. W. Meaningfulness (*m'*) and association value (*a*) in paired-associate syllable learning. *Psychol. Rep.*, 1957, **3**, 441–452.
12. ROCKLYN, E. H., HESSERT, R. B., & BRAUN, H. W. Calibrated materials for verbal learning with middle- and old-aged subjects. *Amer. J. Psychol.*, 1957, **70**, 628–630.
13. SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished Ph.D. dissertation, Yale Univer., 1946.
14. UMEMOTO, T., MORIKAWA, Y., & IBUKI, M. The non-association values and meaningfulness of 1892 Japanese two-letter syllables and words. *Jap. J. Psychol.*, 1955, **26**, 148–155.
15. WICKENS, D. D., & MEYER, D. R. *Psychology.* New York: Dryden, 1955.
16. WINZEN, K. Die Abhängigkeit der paarweisen Assoziation von der Stellung des besser haftenden Gliedes. *Z. Psychol.*, 1921, **86**, 236–252.

(Received July 25, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

SOME VARIABLES AFFECTING LATENT EXTINCTION [1]

ARTHUR R. THOMAS [2]

Northwestern University

A problem of considerable importance to extinction theory in recent years has been that of latent or nonresponse extinction. The experimental paradigm for latent extinction has typically involved two groups. Both groups, experimental and control, learn a task to an equivalent degree. The experimental group is given one or more pre-extinction trials in which *S* is introduced into the previously rewarding situation (e.g., the goal box of a runway) without making the response leading to reward, and finds reward no longer present. A control group is given an equivalent series of trials in a discriminably different situation. Latent extinction exists to the extent that performance decrement in subsequent response extinction trials is significantly more rapid for experimental *S*s than for control *S*s. Thus the essence of latent extinction is that response decrement may result, in part at least, without the response having been made.

Of the 12 studies with rats known to the present author, 10 have confirmed the existence of latent extinction (**2, 3, 4, 8, 9, 10, 11, 12, 14, 16**), while 2 have failed to demonstrate it (**1, 15**). Although the evidence for this phenomenon appears convincing, findings of several of the studies must be accepted with extreme caution. Bugelski, Coyer, and Rogers (**1**) have challenged the positive results of the Seward and Levy study (**16**) on the grounds that time spent on the goal platform during response extinction trials was not equated for experimental and control groups. Other studies confirming the existence of latent extinction have failed to give any sort of pre-extinction experience to control *S*s (**4, 11, 12**). Finally, Scharlock's negative findings (**15**) may have been due to the paucity of cues which could be used to differentiate between the correct and incorrect goal platforms. The use of more easily discriminable goals might have changed the outcome of the experiment.

While it is generally agreed that the Hullian notion of reactive inhibition (**7**) cannot explain latent extinction findings, many alternative theoretical interpretations have been offered. Seward and Levy (**16**), applying Tolman's theory to their results, concluded that the pre-extinction trials which their experimental group received facilitated formation of the expectancy that the goal platform was a "place of nonreward" rather than a "place of reward." Davis (**3**) has explained his positive results in terms of frustration-produced drive, a concept originated by Sheffield (**17**) and Stanley (**18**). More recently, Moltz (**11**) and Moltz and Maddi (**12**) have demonstrated that pre-extinction experience reduces the secondary reward value of cues in the goal box and hence

[1] This article is based upon a dissertation submitted to the Department of Psychology of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to Donald J. Lewis for his encouragement and criticism during the course of the study and to the National Science Foundation, Grant 1060, for financial assistance in purchasing animals for the experimental work.

[2] Now at West Virginia University.

leads to a more rapid performance decrement in subsequent response extinction trials. Each of these theoretical approaches has its merits, but none appears adequate to account for all the facts of latent extinction.

The present study represents a further attempt to verify the existence of latent extinction, and, if it is found to exist, to relate it to Glanzer's (6) notion of stimulus satiation. While derived initially as a theory of response alternation, stimulus satiation appears to the present writer to be a feasible explanation of some of the phenomena of latent extinction. Its major advantage over the theory of reactive inhibition is its assertion that exposure to a stimulus, rather than repetition of a response to the stimulus, is sufficient to produce a decrement in behavior with regard to that stimulus. Specifically, Glanzer states that, with continued exposure to a stimulus situation, amount of satiation to that stimulus and to similar stimuli increases as a negatively accelerated function of time of exposure. Decay of satiation following removal from the stimulus obeys a negatively accelerated decreasing function.

Among the deductions which follow from stimulus satiation are: (*a*) the efficacy of pre-extinction trials in producing a response decrement in extinction varies inversely with the time interval separating the end of pre-extinction training from the beginning of experimental extinction; (*b*) the efficacy of pre-extinction trials in producing a response decrement in extinction varies directly with the length of the pre-extinction trials. The present study, then, attempts not only to demonstrate latent extinction, but to establish an analogy between the response decrement by which it is accompanied and the response decrement due to stimulus satiation. More empirically stated, the present study attempts to determine both the direct and interactive effects upon latent extinction of two stimulus variables: (*a*) the length of the pre-extinction trial, and (*b*) the interval between the pre-extinction trial and the extinction tests.

Method

Apparatus.—A single unit elevated T maze of 2 × 4-in. white pine was employed. The narrow side, about 1¼ in. wide, was used as the runway. The starting platform and stem of the maze were painted gray. The right goal arm and goal platform were painted white; the left, black. Manually operated guillotine doors separated the starting and goal platforms from the body of the maze.

The running surface was 46 in. above the floor. The stem of the maze was 4 ft. long and each arm 1 ft. in length. The starting platform was 10 in. long and 8 in. wide, while the goal platforms were 10 in. long and 4 in. wide. Small 1½ in. diameter china food cups were recessed into the far end of either goal platform with the lip of the cup flush with the surface of the platform. As an added cue, wooden strips ¼ in. wide and ¼ in. high were fastened to the floor of the left goal platform at 1-in. intervals, affording *S* an uneven footing. Coarse sandpaper covered the floor of the right goal platform.

During pre-extinction, a clear plastic shield detained *S* in the goal platform. The shield was mounted on the door frame and the guillotine door was raised to allow *S* maximal visual cues. Control *S*s were given their pre-extinction experience on a 46 in. high, 10 × 6-in. platform of unpainted white pine situated in the corner of the room most remote from the goal platforms.

The only illumination was furnished by a shaded 40-w. lamp in a ceiling fixture above the stem of the maze. In order to maximize the probability of obtaining latent extinction, no attempt was made to control extramaze cues. Differential visual stimulation may have been furnished by a table and a bookcase located on opposite sides of the maze, and differential auditory cues supplied by a Kodak timer used to control intertrial interval.

Subjects.—The *S*s were 90 experimentally naive male albino rats of the Sprague-Dawley strain. On the acquisition day, ages of the *S*s ranged from 61 to 89 days, with a mean of 71 days.

Preliminary procedure.—All *S*s were placed on a maintenance schedule at least seven days before acquisition. A daily ration of 10 gm. of Purina Laboratory Chow was given, and any food remaining after 1 hr. was removed from the cage. Thus *S*s were approximately 22 hr. de-

prived at the beginning of each experimental session. Water was constantly available.

Animals were tamed by *E* for a 10-min. period on each of four successive days. During taming, 45-mg. Noyes Lab Chow pellets identical to those used during acquisiton were presented and all *S*s learned to eat them readily. Following taming, *S*s were given three days of preliminary training in the T maze. On each day, four trials were given. Free choice was allowed on the first two trials, while the third and fourth trials were forced, if necessary, to give *S* two reinforcements daily at each goal. All runs were reinforced with two 45-mg. pellets, and *S* was retained in the goal until the pellets were discovered and eaten. If *S* failed to run from the start platform within a 5-min. period, *E* placed him on one of the goal platforms, determined at random, and allowed him to eat. All *S*s were discarded that failed to run within the 5-min. period of the last pretraining day. Two such *S*s were rejected and were replaced by other *S*s.

Acquisition.—On the day following pretraining, *S*s were given a series of 20 acquisition trials in which food was found only on the nonpreferred side, as determined by the pretraining series. Again, reinforcement consisted of two 45-mg. pellets. The noncorrection method was used and running time and direction of goal turn were measured. Running time was taken from opening of the starting door to entrance into the goal and was read from a stop watch. Direction of goal turn refers to the location of the goal platform into which *S* first entered sufficiently far so that the guillotine door might be lowered. After every trial, *S* was detained in the goal for 10 sec. regardless of the direction of the choice. Intertrial interval was 30 sec., during which time *S* was detained in his home cage with water available. To begin a new trial, *S* was placed on the starting platform facing the door, which was raised after 1 sec.

All experimental groups received the same acquisition training. To assure that all *S*s had adequately learned the habit, two learning criteria were employed: (*a*) a median running time less than 25 sec. on the last five trials; (*b*) at least 9 correct goal choices in the last 10 trials. Seven *S*s were eliminated by failure to meet one or both of these criteria. All were replaced. No *S*s were rejected following the acquisition series.

Pre-extinction.—Pre-extinction experience was given 24 hr. following acquisition. Because of the hypothesized nature of stimulus satiation, it was decided to give only one pre-extinction trial and thereby avoid problems of differential build-up and dissipation of satiation in the various groups. Experimental *S*s were confined in the previously rewarding goal platform with food absent, and control *S*s were placed on the neutral platform. In an attempt to get some measure of the evocation of fractional anticipatory goal responses, number of food cup responses during pre-extinction was recorded for each *S*. Such a response was defined as *S*'s inserting his nose into the food cup below the level of the lip of the cup.

Differential treatment was given to each *S* according to the group to which he was assigned. To test the main and interactive effects of length of pre-extinction period and length of interval between pre-extinction and extinction, a factorial design was used. There were three durations of pre-extinction (1, 5, and 15 min.) and three intervals between pre-extinction and extinction (15 sec., 5 min., and 15 min.). A total of 45 *S*s was used in this design, five *S*s being assigned to each of the nine cells.

To test for the existence of latent extinction effects, 10 control *S*s were run in each of three cells (1-min. pre-extinction trial, 15-sec. delay of extinction; 5-min. trial, 5-min. delay; 15-min. trial, 15-min. delay). An additional five *S*s were run in the experimental cells corresponding to the control cells. Thus main effects of the variables were tested with the 45 experimental *S*s in the factorial design, hereafter called the 3×3 analysis, while the existence of latent extinction was tested with the 60 *S*s (30 experimental and 30 control) in the three diagonal cells and included in the 2×3 analysis. Animals were assigned randomly to conditions, with the restriction that one replication was completed before the next was begun.

Extinction.—After the appropriate delay following pre-extinction, all *S*s were removed from the detention cage and given an identical series of 30 extinction trials during which no food reward was given. As during acquisition, the intertrial interval was 30 sec. and running time and direction of goal turn were recorded. If *S* had not entered a goal platform within a 3-min. period, the trial was terminated and *S* was removed from the maze. Goal platform confinement was 10 sec. following each trial on which a choice was made.

Spontaneous recovery.—Following extinction by a 24-hr. interval, 10 spontaneous recovery trials were given. Response measures, spacing of trials, and other aspects of the procedure were identical to those of original extinction. No reward was present.

Results

Acquisition

The median running time for the last five acquisition trials was obtained for each *S*, and analysis of variance of log median running time

was performed. No significant differences between groups were found either for this measure or for number of correct turns.

Extinction

Running time measures.—Due to extreme skewness, it was impossible to transform extinction scores to meet the assumptions of normality and homogeneity of variance required for the *F* test. In its place, a distribution-free analysis presented by Mood (13) was used for both running time and correct turn data. This analysis is based on ranking of scores and has a chi-square distribution.

For analysis as well as for graphic presentation, the 30 extinction trials for each *S* were broken down into six blocks of five trials each. The median running time for each block was found and the log of this value was used for analysis. All trials terminated by *E* at the end of 3 min. were considered as 180-sec. latencies.

Two running time measures were analyzed: (*a*) raw scores on Trial 1, and (*b*) log median running time for all six 5-trial blocks during extinction. These measures are shown graphically in Fig. 1A, 1B, and 2A, Fig. 1 being based upon the 45 *Ss* in the 3 × 3 analysis and Fig. 2A upon the 60 *Ss* in the 2 × 3 analysis.

The 2 × 3 analysis for Trial 1 scores showed a significantly slower running time for experimental *Ss* than for control *Ss* ($\chi^2 = 4.053$, $P < .05$), while a similar analysis based upon all 30 trials was not significant. Thus, as reflected in Fig. 2A, the effects of pre-extinction experience upon running time were initially large, but transitory.[3]

The main effects of the two independent variables upon running time were assessed by 3 × 3 analyses, and are shown graphically in Fig. 1.

[3] An additional analysis based upon the first 15 extinction trials showed a significant difference between experimental and control groups.

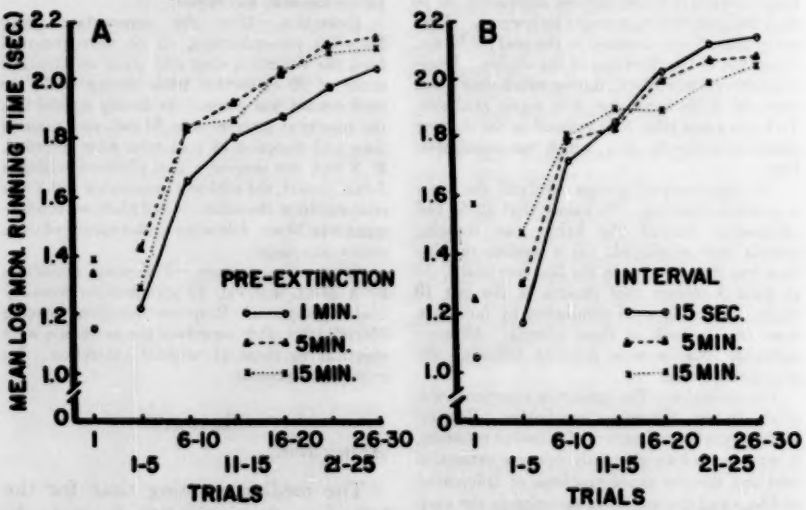

FIG. 1. A. Running time during extinction for various pre-extinction times. B. Running time during extinction for various intervals between pre-extinction and extinction.

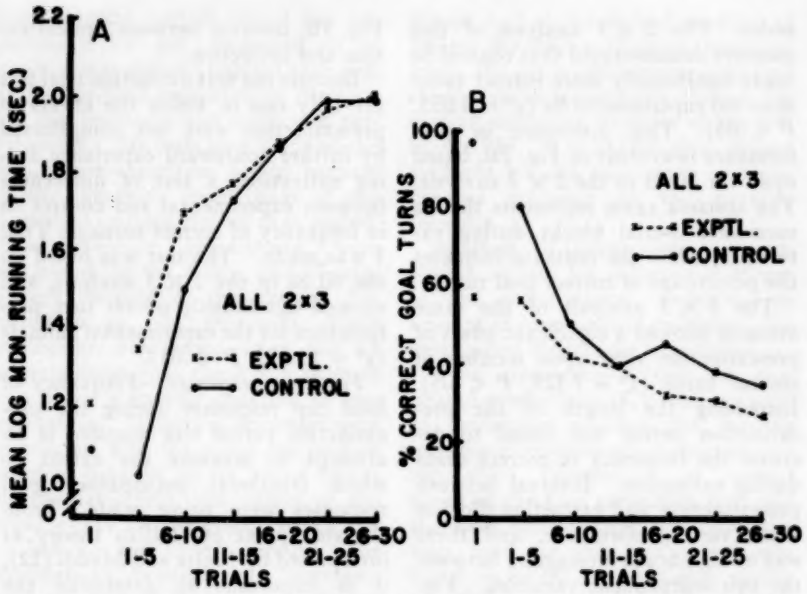

FIG. 2. A. Running time during extinction for all experimental and control animals in the 2 × 3 analysis. B. Correct turns during extinction for all experimental and control animals in the 2 × 3 analysis.

For Trial 1 only, as interval between pre-extinction and extinction increased, running time increased significantly ($\chi^2 = 12.500$, $P < .01$). This effect was not evident when all 30 trials were studied. No significant effects were found either for the second independent variable, length of the pre-extinction trial, or for the interaction between the two variables.[4]

A final significant finding is evident upon inspection of Fig. 1B. Animals with the 15-sec. interval between pre-extinction and extinction run most rapidly on the initial extinction trials, while they are the slowest of the three groups by the end of the extinction series. To test the significance of the interaction between extinction trial and interval between pre-extinction and extinction, Edwards' analysis of variance for repeated measurements on the same *Ss* (**5**, p. 288) was employed. Log median running times on extinction Trials 1–5 and 26–30 for the 45 experimental *Ss* in the 3 × 3 design were included in the analysis. While the assumption of normality was not met, the *F* ratio for interaction was significant at well beyond the 1% level ($F = 34.21$, $df = 2,42$).

Turn measures.—Mood's distribution-free analysis was also applied to turn data during extinction. The principal measure studied was the number of correct goal turns made by each *S* during the 30-trial extinction

[4] It must be noted that Mood's test for interaction invariably results in a negative chi square. As a further check on the significance of interaction effects, normal curve analyses were also performed. While not strictly appropriate because the data violated the assumption of normality, such tests consistently yielded results in accord with those of the Mood test.

series. The 2 × 3 analysis of this measure demonstrated that control *Ss* made significantly more correct turns than did experimental *Ss* ($\chi^2 = 4.053$, $P < .05$). This difference in performance is evident in Fig. 2B, based upon the 60 *Ss* in the 2 × 3 analysis. The abscissa again represents the six successive 5-trial blocks during extinction, while the ordinate indicates the percentage of correct goal turns.

The 3 × 3 analysis of the same measure showed a significant effect of pre-extinction time upon number of correct turns ($\chi^2 = 7.125$, $P < .05$). Increasing the length of the pre-extinction period was found to decrease the frequency of correct turns during extinction. Interval between pre-extinction and extinction did not affect turn performance, and there was no significant interaction between the two independent variables. Figure 3, based upon the 45 experimental *Ss* in the 3 × 3 analysis, portrays these results. In Fig. 3A, length of pre-extinction is the parameter; in Fig. 3B, interval between pre-extinction and extinction.

Because the first extinction trial was the only one in which the effects of pre-extinction were not complicated by further nonreward experience during extinction, a test of differences between experimental and control *Ss* in frequency of correct turns on Trial 1 was made. The test was based on the 60 *Ss* in the 2 × 3 analysis, and showed significantly poorer turn performance for the experimental animals ($\chi^2 = 11.273$, $P < .01$).

Food cup responses.—Frequency of food cup responses during the pre-extinction period was recorded in an attempt to measure the extent to which fractional anticipatory goal responses were being made. From the standpoint of Hullian theory as interpreted by Moltz and Maddi (12), it is important to determine the relationship between such responses and subsequent extinction performance.

Because varying the length of the

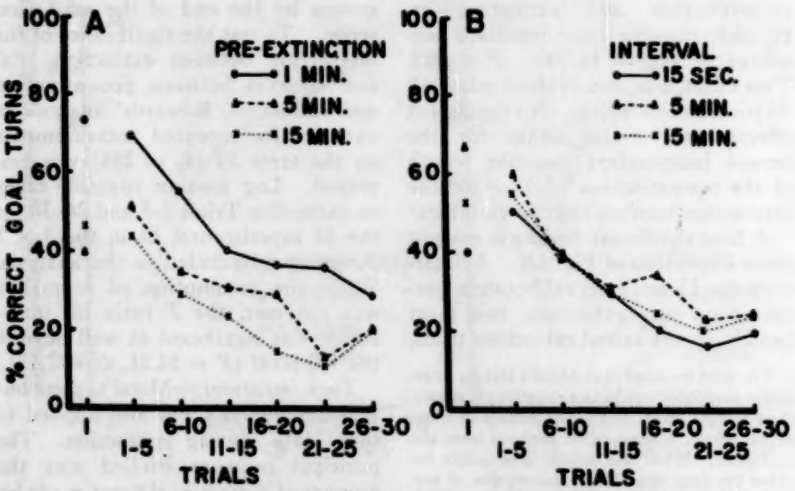

FIG. 3. A. Correct turns during extinction for various pre-extinction times. B. Correct turns during extinction for various intervals between pre-extinction and extinction.

pre-extinction period varied the opportunity for cup responses to be made, it was decided to treat each pre-extinction time separately in correlating cup responses and extinction performance. No tests were made on the 1-min. pre-extinction group because of the extremely small range of cup responses emitted during that period. Rank-difference correlations were computed for the 5-min. and 15-min. groups because of the skewed distribution of cup responses. Each coefficient was based upon the 15 experimental *Ss* in the 3 × 3 analysis run under a specific duration of pre-extinction. When cup responses were correlated with the mean log median running time for all six 5-trial blocks during extinction, the values of rho were −.414 for the 5-min. group and −.303 for the 15-min. group. The correlations between cup responses and number of correct turns during the 30 extinction trials were .288 and .499 for the 5-min. and 15-min. groups, respectively. The value of rho needed for two-tailed significance at the 5% level is .524. Thus, although not reaching statistical significance, the signs of the coefficients indicate that *Ss* making more cup responses tend to run more rapidly and make more correct turns in extinction than *Ss* investigating the food cup less often.

Spontaneous Recovery

Running time measures.—To test for spontaneous recovery of the running response, the Wilcoxon signed rank test for paired observations (**19,** p. 432) was used. This analysis was based on the 90 difference scores, one from each *S*, obtained by subtracting log running time on the first spontaneous recovery trial from log mean running time for the last block of five extinction trials (Trials 26–30). Since almost half of the *Ss* had medians of 180 sec. for the last block of extinction trials, the log mean running time was chosen as more sensitive than a log median measure. Results of the test ($z = 6.915$, $P < .0001$) give strong evidence for the occurrence of spontaneous recovery of the running response.

Analysis of covariance was employed to test for differential effects of the experimental variables on amount of spontaneous recovery. No meaningful differences were found.

Turn measures.—A test for correlated proportions (**5,** p. 86) was used to determine whether the turning response gave evidence of spontaneous recovery. Frequencies of correct turns on the final extinction trial and the first spontaneous recovery trial were used for this test, which was based on all 90 *Ss*. A highly significant difference was found ($z = 3.817$, $P < .0001$) in a direction favorable to the existence of spontaneous recovery. As in the case of the running response, amount of recovery of the turning response was not significantly influenced by any of the experimental variables.

Discussion

Results of this study clearly support the existence of latent extinction. Experimental *Ss* run more slowly and make fewer correct responses than corresponding control *Ss*. Experimental vs. control differences in running time are significant only for the first part of the extinction series, while differences in correct turns persist throughout all 30 trials.

However, the data fail to give unequivocal confirmation to the specific predictions made from Glanzer's (**6**) theory of stimulus satiation. One of the predictions—namely, that amount of performance decrement in extinction is directly related to length of the pre-extinction period—was confirmed with the correct turn data in Fig. 3A. Pre-

extinction duration had no effect upon the running time measure. No support was found for the prediction that the performance decrement of latent extinction would decrease as the interval between pre-extinction and extinction increased. In fact, the opposite relationship was found for the running response on the first extinction trial: a shorter interval was followed by faster running time than was a longer interval. Interval and direction of turn were not significantly related.

Thus the empirical findings partially support one of the two predictions from stimulus satiation and partially refute the other. Not only is some of the evidence directly contradictory to Glanzer's theory, but the theory cannot account for the differential effects of the two independent variables upon the two response measures. Alternative explanations of the data might prove more profitable.

Expectancy theory, of course, has no trouble in handling the phenomenon of latent extinction. However, the lack of data on the temporal factors involved in the acquisition and extinction of expectancies precludes any specific predictions for this experiment.

The present results appear relevant to the Hullian approach adopted by Moltz and Maddi (12). From the findings of these investigators, it seems legitimate to predict that, the greater the degree of extinction of secondary reward value of the goal platform cues during pre-extinction the greater the performance decrement during extinction. It is here necessary to assume that the fractional anticipatory goal responses (i.e., cup responses) recorded during the pre-extinction period bear some relation to extinction of secondary reward. It will be remembered that correlations between food cup responses and running time during extinction were negative, and those between cup responses and correct turns were positive. If one makes the assumption that *S*s investigate the food cup more often when its secondary reward value is high than when it is low, these results can readily be reconciled with those of Moltz and Maddi. That is, those *S*s making many food cup responses during pre-extinction do so because the secondary reward value of the food cup has not diminished appreciably, and such *S*s show the least performance decrement during extinction.

Also consistent with this approach is the finding that a significant and inverse relationship exists between length of the pre-extinction period and frequency of correct turns during extinction. More extinction of secondary reward should occur within a 15-min. pre-extinction interval than a 1-min. interval, and, consequently, a greater performance decrement during extinction should ensue.

In spite of the agreement between the present data and the approach of Moltz and Maddi, there still remains the question of why the two response variables of the present study were differentially affected by the two stimulus variables. Davis' (3) findings appear pertinent at this point. Using a T maze, Davis found that his experimental *S*s, which received pre-extinction experience, made fewer correct turns during extinction than controls, but actually ran faster than the controls. Davis has used the notion of frustration-produced drive to explain his results. Following the argument of Sheffield (17) and Stanley (18), he has suggested that pre-extinction confinement may increase *S*'s drive level and thus facilitate the running response, for which few competing responses exist. The number of correct turns does not increase because a competing response (i.e., an incorrect turn) is readily available.

It is not the intention of the present author to make a detailed application of the theory of frustration-produced drive to the results of this study. The approach, however, has the virtue of recognizing, even if not predicting accurately, differential effects of pre-extinction experience on different response measures. Regrettably, little work has been done to establish the temporal course of build-up and dissipation of

frustration-produced drive. With certain auxiliary assumptions concerning these temporal factors, nearly all the data of the present study may be incorporated into the frustration-produced drive approach. For example, if it is assumed that a substantial amount of such drive dissipates in a 15-min. rest interval (the intertrial interval for the spaced extinction groups in the Sheffield and Stanley studies), it is easy to explain the finding of significantly slower running after a 15-min. interval between pre-extinction and extinction than after a 15-sec. interval. Other similar assumptions may be made to explain other findings both in this study and in other studies of latent extinction. Such an approach is quite tentative, however, until more quantitative data on frustration drive exist.

It must be noted that the two response measures, running time and correct turns, cannot be completely independent. That is, *S* cannot make a correct turn without also having run within the 3-min. interval allowed for each extinction trial. Three separate point biserial correlations based upon all 90 *S*s were computed between direction of turn and log running time for extinction Trials 1, 15, and 30. A correct turn was arbitrarily assigned a value of 1 and an incorrect turn or no turn a value of 0. The correlations, −.38, −.30, and −.52, respectively, are all significant at the 1% level and indicate a longer running time associated with incorrect than with correct turns. However, the correlations are not of such magnitude as to prevent differential effects of the two independent variables on the response measures.

Summary

The present research tests (*a*) the existence of the phenomenon of latent extinction, and (*b*) the effects of length of pre-extinction period and length of interval between pre-extinction and extinction upon the amount of latent extinction. Ninety rats were given 20 acquisition trials on a single-unit elevated T maze with food reward on the nonpreferred side. Differential treatment for the various groups was instituted on the extinction day, 24 hr. following acquisition. Experimental *S*s were given pre-extinction experience by being confined in the previously rewarding goal platform with no food present, while control *S*s were confined for a similar length of time on a neutral platform. One such trial was given, its length being 1, 5, or 15 min. The interval between pre-extinction and extinction was 15 sec., 5 min., or 15 min. Thirty response extinction trials were given to all *S*s, and running time and correct turns recorded. Ten nonrewarded spontaneous recovery trials followed extinction by 24 hr.

A 3 × 3 factorial design based on 45 *S*s was used to evaluate the effects of the independent variables upon latent extinction, while its existence was determined by comparing performance of 30 experimental and 30 control *S*s. Comparisons of experimental and control groups showed a significant depression in number of correct turns for experimental *S*s lasting throughout the extinction series. Running time was increased, but differences were significant only for the first part of the extinction series.

It was found that increasing the length of the pre-extinction period significantly reduced number of correct turns during the extinction series. Increasing the length of the interval between pre-extinction and extinction significantly increased running time on the first extinction trial. Length of the pre-extinction period was not related to running time, nor was interval between pre-extinction and extinction related to number of correct turns. Significant spontaneous recovery occurred for both running and turning responses. No meaningful differences between groups in amount of recovery were found, however.

The existence of latent extinction was upheld by both response measures, and both independent variables were found to have significant effects. Results are best explained by the notion of frustration-produced drive.

References

1. Bugelski, B. R., Coyer, R. A., & Rogers, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.*, 1952, **44**, 27–30.
2. Coate, W. B. Weakening of conditioned bar-pressing by prior extinction of its subsequent discriminated operant. *J. comp. physiol. Psychol.*, 1956, **49**, 135–138.
3. Davis, J. D. Extinction as a function of exposure to an empty goal box. Unpublished honors thesis, Brown Univer., 1954.
4. Deese, J. The extinction of a discrimination without performance of the choice response. *J. comp. physiol. Psychol.*, 1951, **44**, 362–366.

5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Glanzer, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.
7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
8. Hurwitz, J. M. B. Response elimination without performance. *Quart. J. exp. Psychol.*, 1955, **7**, 1–7.
9. Levy, N. Latent extinction of a jumping response. *Amer. Psychologist*, 1955, **10**, 416–417. (Abstract)
10. Lewis, D. J., & Cotton, J. W. Partial reinforcement and nonresponse acquisition. *J. comp. physiol. Psychol.*, 1958, **51**, 251–254.
11. Moltz, H. Latent extinction and the reduction of secondary reward value. *J. exp. Psychol.*, 1955, **49**, 395–400.
12. Moltz, H. & Maddi, S. R. Reduction of secondary reward value as a function of drive strength during latent extinction. *J. exp. Psychol.*, 1956, **52**, 71–76.
13. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
14. Pliskoff, S. S. Response elimination as a function of generalization, motivation, and number of non-response extinction trials. Unpublished doctor's dissertation, New York Univer., 1955.
15. Scharlock, D. P. The effects of a pre-extinction procedure on the extinction of place and response performance in a T maze. *J. exp. Psychol.*, 1954, **48**, 31–36.
16. Seward, J. P., & Levy, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660–668.
17. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
18. Stanley, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.
19. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

(Received August 18, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

STUDIES OF DISTRIBUTED PRACTICE: XVIII. THE INFLUENCE OF MEANINGFULNESS AND INTRALIST SIMILARITY OF SERIAL NONSENSE LISTS[1,2]

BENTON J. UNDERWOOD AND JACK RICHARDSON
Northwestern University *Harpur College*

The primary purpose of the present experiment is to determine if intertrial interval interacts with meaningfulness and with intralist similarity in the learning of serial nonsense lists. It was undertaken in the belief that no available data give uncontaminated tests of these two interactions. The following discussion gives the bases for this belief.

Certain lines of evidence could be advanced in support of the principle that the lower the meaningfulness of the material the greater the likelihood that distributed practice will facilitate learning.

(*a*) Tsao (**3**) has pointed out that the older studies using prose materials commonly found that massed practice was as good or better than distributed practice for learning, whereas with syllables the opposite was usually true. Tsao also points out, however, that one cannot be sure that a difference in meaningfulness is the only critical difference between prose material and nonsense syllables.

(*b*) In two previous reports in this series (**4**, **5**), both of which were performed under identical conditions except that serial adjectives were used in one case and syllables in the other, distributed practice appeared to exert a greater facilitating effect on the learning of syllables than on the learning of adjectives. However, since it is now known that intralist similarity and meaningfulness interact in that the lower the meaningfulness the greater the influence of a given degree of intralist similarity (**9**), the difference in the effect of intertrial interval in the two experiments may be due to differences in intralist similarity and not meaningfulness.

(*c*) The one direct attack on the problem was made by Tsao (**3**). Two serial lists of syllables of high meaningfulness were constructed and two with low meaningfulness. One list from each level was presented by massed practice and one by distributed practice, the latter consisting of a 1-min. rest after every two trials. All lists were presented for 10 trials and all *Ss* served in all conditions by counterbalancing. The low-meaningful lists were significantly facilitated by distributed practice, the high-meaningful lists were not. Thus, if Tsao had performed an analysis of variance on his data it would be expected that the interaction term (Intertrial interval × Meaningfulness) would have been significant. Yet, two facts about Tsao's study make it seem desirable to have further work on this interaction phenomenon. First, an examination of Tsao's lists shows that the low-meaningful lists have greater intralist similarity (fewer different letters used) than the high-meaningful lists. Thus, meaningfulness and intralist similarity may be confounded. Second, all *Ss* served in all conditions. It is now known that such designs may introduce considerable interference between lists and this interference may well be different for different levels of meaningfulness. Thus, it is possible that Tsao's results might be due to differences in intralist or interlist similarity, or both.

The primary implication of the above comments is that intralist similarity may have confounded tests of the interaction between meaningfulness and intertrial interval. Such a criticism is not valid, of course, unless intralist similarity and intertrial interval do indeed interact in learning. In two studies (**4**, **5**) in which intertrial interval has been varied, intralist similarity has also been manipulated. No interaction was apparent. But, neither of these studies was a satisfactory test of the interaction because they were counter-

[1] R. W. Schulz supervised the gathering of the data and has critically read the manuscript.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

balanced designs which are now known (7) to result in a reciprocal relationship between intralist and interlist similarity. Therefore, if distributed practice facilitates learning when interference is high (produced either by intralist or interlist similarity) the two experiments are simply not adequate tests.

It appears, therefore, that both meaningfulness and intralist similarity must be manipulated independently. Interlist similarity can be eliminated by using *Ss* naive to verbal learning who serve in only one condition of the experiment. The present study attempts to meet these specifications.

Method

Materials.—Four serial lists of 10 syllables each were used. These lists have been used in a previous study and are reproduced there (9). Two lists were made up of Glaze (1) syllables of from 0.0% to 20.0% association value, and two with syllables of from 93.3% to 100% association values. One of the lists at each level of meaningfulness had high intralist similarity and one low intralist similarity. The two lists with high intralist similarity had only four consonants, each being used five times. In the lists with low intralist similarity each consonant was used only once. All vowels were used twice in each list. The four lists are designated as follows: HM-HS (high meaningfulness, high intralist similarity); HM-LS (high meaningfulness, low intralist similarity); LM-HS (low meaningfulness, high intralist similarity), and LM-LS (low meaningfulness, low intralist similarity). (In the previous publication, where the lists are reproduced, the fourth syllable in the HM-LS list was given as MEK; this should have been MEX).

Each list was preceded by three asterisks as the anticipatory cue for the first syllable. As in the previous study, 10 different serial orders of each list were used so that each syllable occurred equally often at each serial position. The rate of presentation was 2 sec. and *S* spelled the syllables, using the anticipation method.

Subjects and procedure.—Eight groups of 50 *Ss* were used. Two groups were assigned to each list, one learning by massed practice and one by distributed practice. Massed practice consisted of an 8-sec. blank period between each trial. Distributed practice consisted of 38 sec. between each trial, this interval being filled with symbol cancellation. With two levels of meaningfulness, two levels of intralist similarity, and two intertrial intervals, the experiment becomes a $2 \times 2 \times 2$ design with independent groups in each cell.

A random order of the eight conditions (four lists, with massing and distribution for each) was made up subject to the restriction that each condition occur 50 times. The *Ss* were then assigned in succession in order of their appearance at the laboratory. When an *S* was dropped for any reason he was replaced with the next *S*. A total of 35 *Ss* was dropped for a variety of reasons. No appreciable bias can result from this since the minimum number dropped from any one condition was 2 and the maximum 7. To avoid bias in selection of *Ss* due to differential difficulty produced by the different conditions, a constant number of trials was used for all lists. Under this procedure only two *Ss* were dropped for failure to learn. One *S* became so emotionally upset in learning that it seemed advisable to dismiss her. The other *S*, when he experienced difficulty in learning, simply quit and walked out of the room. All *Ss* used in the present evaluation were naive to laboratory verbal learning experiments.

Each list was always presented for 30 anticipation trials with *S* instructed to give as many correct responses as possible on each trial. The *S* returned for five relearning trials 24 hr. after learning.

Results

Learning.—The basic response measure used was the mean number of correct responses over the 30 trials. Table 1 shows the means for each condition and Table 2 shows the summary of the analysis of variance.

As in the previous study using these lists (9), meaningfulness and intralist similarity produced highly significant

TABLE 1

Mean Number of Correct Responses for Trials 1–30

List	Massed*	Distributed*
HM-HS	157.12	188.86
HM-LS	206.42	203.70
LM-HS	110.64	123.62
LM-LS	165.46	176.38

* Estimate of σ_M, based on the within-groups variance, is 6.08.

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RESPONSES IN TRIALS 1–30

Source	*df*	*MS*	*F**
Interval (I)	1	17,503.54	9.48
Meaningfulness (M)	1	202,500.25	109.65
Similarity (S)	1	184,298.74	99.80
I × M	1	163.84	—
I × S	1	8,335.69	4.51
M × S	1	11,793.96	6.39
I × M × S	1	6,560.25	3.55
Within	392	1,846.75	
Total	399		

* With 1 and 400 *df*, *F* at .05 level is 3.86, and at .01, 6.70.

effects. The interaction between these two variables is also significant, i.e., the lower the meaningfulness the greater the effect of intralist similarity.

Turning next to the major data, with special reference to the intertrial-interval interactions, on the basis of the statistical evidence the following conclusions would be reached: (*a*) over all lists, distributed practice is superior to massed practice beyond the 1% level of significance; (*b*) the interaction between meaningfulness and intertrial interval does not even approach significance; (*c*) interaction between intertrial interval and intralist similarity is significant between the 1% and 5% levels, and (*d*) the triple interaction approaches the 5% level. While these conclusions cannot be gainsaid, they do not fully impart the effects produced by the variables The complete picture can best be perceived through an inspection of the acquisition curves (Fig. 1) together with the means in Table 1.

Considering first the effect of distributed practice, it can be seen that the greatest effect was produced in learning List HM-HS (this is reflected in the triple interaction which approached the 5% level of significance). Indeed, when *t* ratios were calculated between massing and distribution for each of the two low-meaningful lists the null hypothesis could not be rejected in either case. However, it would not seem appropriate to conclude that intertrial interval had no influence in learning these two lists. After approximately the tenth trial, distributed practice is consistently superior to massing. Furthermore, the positive effects of distribution for these two lists were of sufficient magnitude that the *F* for the meaningfulness-intertrial interval interaction was not significant. This is to say that the combined effects of distributed practice for these two low-meaningful lists were of the same order of magnitude as the effects for the single high-meaningful list (HM-HS). Another way of viewing the intertrial interval-meaningfulness interaction is to note that with high-similarity lists distribution favors the high-meaningful lists, but with low similarity, distribution favors the low-meaningful lists. By any way of looking at the data, it must be concluded that within the range of meaningfulness used here, there is no consistent interaction between meaningfulness and intertrial interval. However, as will be pointed out later, the data suggest certain effects of meaningfulness which may be relevant for interpreting the effects of distributed practice.

Turning to intralist similarity, it is apparent that for lists of high meaningfulness this variable is critical in determining whether or not distribution facilitates learning. But, this principle is not easily supported for the low-meaningful lists. However, the level of learning attained in 30 trials for List LM-LS is appreciably higher than for List LM-HS. It might be suggested, therefore, that had the *S*s learning List LM-HS continued until they equaled the performance achieved by those *S*s learning List LM-LS, the effect of distribution would have been shown to be greater for the high-similarity list. An argument opposed to this would be to note that the curves for List LM-HS do not indicate a gradually widening difference between

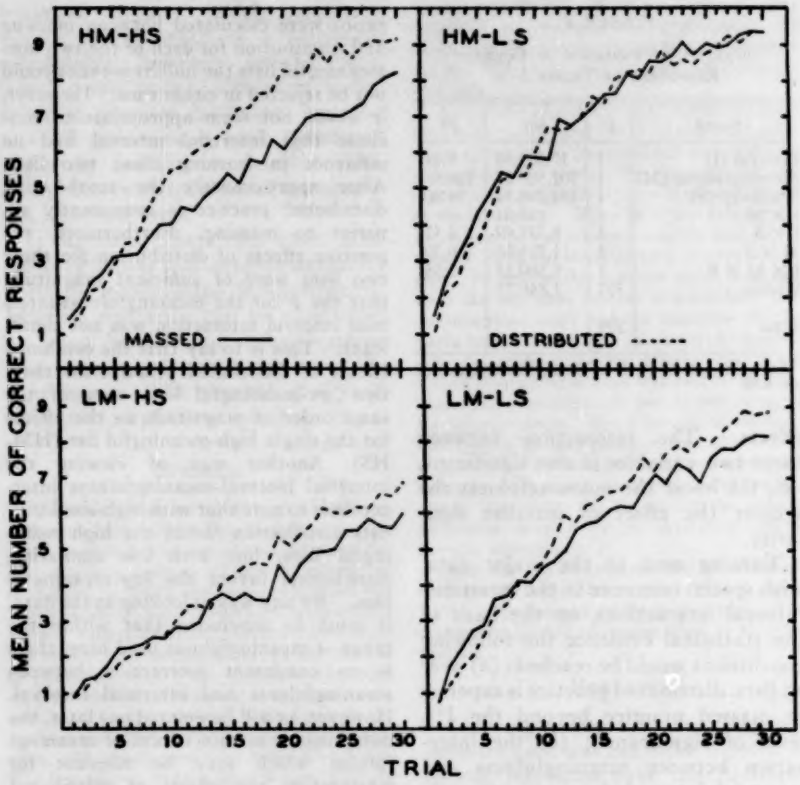

FIG. 1. Acquisition curves for each list by massed and distributed practice.

massing and distribution in the latter part of learning. On the basis of the data as shown, therefore, it appears that intralist similarity interacts less with intertrial interval for the low-meaningful lists than it does for the high-meaningful lists. This suggests again, of course, that the triple interaction may indicate a reliable set of differences.

One further point about the acquisition curves should be noted. For all four lists the initial performance (for from 5 to 8 trials) is poorer under distributed practice than under massed practice. The effect of this is to reduce the differences between massed and distributed practice when total correct responses over all trials is used as the response measure. This cross-over phenomenon will be discussed in the next section.

Retention.—Due to the fact that the level of learning at the end of 30 trials was quite different for the different lists, the retention measurements primarily reflect these differences in original strength. However, certain facts will be mentioned briefly. First, there is no difference in recall as a function of massing and distribution for any list when adjustments are made for differences in associative strength at end of learning. The

massed and distributed conditions are combined, therefore, for further comparisons.

The high recall under all conditions is worth noting. As pointed out elsewhere (8), if naive *Ss* are used so that no proactive interference from previously learned lists is present, very little loss should occur over 24 hr. To demonstrate this for the present data, the mean number of items correct on the last learning trial has been used as a base for calculating percentage recall. It can be seen from Fig. 1 that little if any error is introduced by using this base. That is, if *Ss* had been given one more trial their performance would have been only slightly higher than it was on Trial 30. As shown in Table 3, percentage recall varies from approximately 72% for the most poorly learned list (LM-HS) to nearly 90% for the strongest (HM-LS).

In confirmation of the previous study using these lists (9), no difference in recall is observable as a function of meaningfulness when adjustments are made for differences in associative strength at the end of learning. Concerning intralist similarity and recall, the two experiments require discussion. In the previous study, more loss was reported for lists of high intralist similarity than for low when both were of low meaningfulness. The difference occurred primarily with items of weak associative strength. That is, weak items in the high-similarity list were forgotten more rapidly than comparably weak items in the low-similarity list. When the present data were examined no such difference was apparent. Two possibilities for the discrepancy seemed possible. First, the difference in the earlier experiment may have been a chance phenomenon. The other possibility was that the method used to compare retention was biased against the high-similarity list. In the previous study a gross probability analysis (6) had been used to equate for associative strength. Since the time that those data were so evaluated, a more precise method of probability analysis has been worked out. It had been discovered that the gross method did produce some bias in some situations. The more refined method avoids this bias. Without detailing this method it is sufficient to say that it makes estimates of associative strength at the end of learning for items by using for this estimation only items which have approximately the same strength as determined by number of times given correctly during learning. In the gross probability analysis this was not true. In the present data the refined method was used and no difference was obtained in recall as a function of intralist similarity at any level of associative strength. Therefore, the data from the earlier experiment were reanalyzed by the newer technique and the difference in recall as a function of intralist similarity was no longer present. The conclusion must be, therefore, that intralist similarity as manipulated here does not influence retention as measured by recall.

TABLE 3

PERCENTAGE RECALL BASED ON NUMBER OF ITEMS CORRECT ON THE LAST LEARNING TRIAL

List	Correct on Last Trial	Correct on Recall	% Recall
HM-HS	8.73	7.12	81.6
HM-LS	9.35	8.39	89.7
LM-HS	6.45	4.63	71.8
LM-LS	8.55	7.06	82.6

In summary of the retention data, the only variable found to influence recall was associative strength. Meaningfulness, intralist similarity and intertrial interval are not relevant variables in the recall of the present lists.

DISCUSSION

The present data leave no doubt that intralist similarity is a highly important variable in determining whether or not

distributed practice facilitates learning. While meaningfulness and intertrial interval did not interact in any simple way, the argument will be advanced that meaningfulness cannot be ignored entirely in considering the effects of distribution. With high-meaningful lists facilitation by distribution occurred only when intralist similarity was high. If meaningfulness is irrelevant, and only intralist similarity relevant, why was not distribution as effective for the list of high intralist similarity and low meaningfulness as it was for the list of high intralist similarity and high meaningfulness (HM-HS vs. LM-HS)?

It has already been pointed out that both in this study and in the previous one using the same lists, meaningfulness and intralist similarity interact. Given a constant degree of intralist similarity as determined by amount of duplication of consonants, learning will be more retarded with low-meaningful lists than with high-meaningful lists. The reason for this, it is believed, lies in the nature of the task presented *S* when given high- and low-meaningful syllables. Most high-meaningful syllables will be perceived as words. This is not true for low-meaningful syllables. In the latter case *S* must learn to integrate (in the Mandler sense [2]) the individual letters into a unitary response. Therefore, it may well be that letter duplication in the high-meaningful syllables will produce less interference than in the low-meaningful syllables. The high-meaningful syllables provide a discriminative cue of "wordness" not frequently present in the low-meaningful syllables. Thus, the same formal degree of intralist similarity could produce more interference among low-meaningful syllables than among high. The argument is, therefore, that interference is a critical variable determining whether or not distribution facilitates. Meaningfulness enters only because a given degree of intralist similarity produces more interference in low-meaningful lists than it does in high-meaningful lists. The fact that distributed practice affected the learning of the low-meaningful list with low intralist similarity (LM-LS) even a small amount could be attributed to the fact that in these low-meaningful lists the duplicated vowels produce an interference which does not occur in the high-meaningful list.

The question still remains, however, as to why List LM-HS did not show greater facilitation under distribution than it actually did. If interference is a critical factor, this list, which produced more interference than any other list, should certainly have been facilitated by distribution of practice more than the other lists. A simple hypothesis that amount of facilitation by distributed practice and amount of interference are directly related cannot account for this discrepancy. Thus, while the present data leave no doubt that interference produced by intralist similarity is a highly important factor determining whether or not distributed practice facilitates learning, some additional factor must be eventually included as a modifying agent.

It was noted that for every list massed practice was initially superior to distributed practice (Fig. 1). The main question concerning this observation is whether or not systematic relevance should be attributed to it. That is, does it have implications for a theory of distributed practice? The evidence, while not conclusive, suggests a negative answer. The *S*s in the distributed-practice group had no previous experience in switching from the learning task to the cancellation task and back. The poorer performance may simply represent the retardation in learning as *S* learned to make the transitions. It might, too, represent slower warm-up than occurs under massed practice. An examination of the serial-position curves shows that the *S*s having distributed practice are poorer on the initial items in the list than are those *S*s having massed practice. If the *S*s having distributed practice were having trouble in switching from cancellation to learning, it would be expected that the initial items in the list would be adversely affected. The fact that the phenomenon occurs in List HM-LS which is not facilitated by distributed

practice also suggests that the initial decrement of the distributed group should not be given theoretical relevance.

Summary

The major purpose of this study was to determine if meaningfulness and intralist similarity interact with intertrial interval in learning serial nonsense lists. Eight groups of 50 *Ss* each learned a single list of 10 serial syllables. The eight cells of the experiment were made up of two levels of meaningfulness (thus, a total of four lists), and two intertrial intervals for each list (massed and distributed practice). Learning was carried for 30 trials for all lists with recall after 24 hr.

The major results were: (*a*) Meaningfulness does not interact with intertrial interval. (*b*) Intralist similarity does interact with intertrial interval but in a complex fashion. With high meaningfulness, high intralist similarity is clearly necessary to produce facilitation by distributed practice. With low meaningfulness, intralist similarity does not influence learning by massed vs. distributed practice. (*c*) Recall was affected only by strength of association.

The results indicate that intralist interference is necessary before distributed practice facilitates serial learning. However, the relationship is not simple; the data do not indicate that the greater the intralist interference the greater the facilitation in learning by distributed practice. Some factor(s) modifies this relationship when a certain degree of interference is reached.

References

1. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
2. Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.
3. Tsao, J. C. Studies in spaced and massed learning: II. Meaningfulness of material and distribution of practice. *Quart. J. exp. Psychol.*, 1948, **1**, 79–84.
4. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
5. Underwood, B. J. Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjective lists. *J. exp. Psychol.*, 1953, **45**, 253–259.
6. Underwood, B. J. Speed of learning and amount retained: A considertion of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.
7. Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 160–166.
8. Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.
9. Underwood, B. J., & Richardson, J. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp. Psychol.*, 1956, **52**, 119–126.

(Received August 29, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

AN INVESTIGATION OF MEDIATION IN PRECONDITIONING [1]

ROBERT J. SEIDEL [2]

University of Pennsylvania

The present study is an investigation into the nature of the mediation process in learning, hypothesized by Osgood (7) and Mowrer (6) to be peripheral in nature. Mowrer specifies autonomic response mediation, while Osgood stresses a general peripheral type. The latter does, however, imply an autonomic basis of the mediation process where plausible. Cognitive theorists (3, 4), on the other hand, have emphasized S-S contiguity as the necessary and sufficient condition for learning. From this standpoint, the autonomic response factor would be unimportant for mediation.

The studies done by Bahrick (1, 2) suggest a possible means of opposing the S-R and S-S theories in a sensory preconditioning (SPC) framework. Bahrick exposed rats under hunger to pairings of buzzer and light without imposing any response requirement. Later he gave them the training-transfer sequence of SPC under satiation (for food). The task learned in this sequence was an avoidance response in a shuttle box. The *Ss* showed a positive transfer effect from one SPC stimulus to the other regardless of which stimulus was used first. This occurred despite the fact that the hunger state, present during the exposure period, was absent.

These data suggest that the autonomic processes, which had been distinctly different between exposure and the training-test sequence, might not have been important in mediating the buzzer-light association. No definitive answer could be obtained from Bahrick's study, however, since there had been no systematic investigation of the autonomic factor in mediation. Such an approach was attempted in the present SPC study by varying the consistency of the autonomic responses dominant during the exposure period (e.g., parasympathetic) with those dominant during the training-testing sequence (e.g., sympathetic).

According to the mediation view noted above, experimental groups in comparison to control *Ss* should exhibit degrees of transfer effect which would be (*a*) *positive* and relative to the degree of consistency of autonomic responses between exposure and the training-testing sequence, or (*b*) *negative* and relative to the degree of active opposition of the drive states present during the training-testing sequence. The general prediction of *negative* transfer effect stems from the high probability that the drive state during exposure (hunger) and that during the training-testing sequence (fear) were autonomically opposed.

METHOD

Subjects

The *Ss* were 41 hooded male rats from the Rockland Farm stock. They weighed approxi-

[1] This paper is a portion of a dissertation submitted to the Department of Psychology of the University of Pennsylvania in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his gratitude to William A. Shaw for his persistent encouragement and guidance throughout the development of this study. The writer is also indebted to Ronald H. Forgus and Seymour Feshbach for their invaluable advice during the course of this investigation.

[2] Now at Air Force Special Weapons Center, Kirtland Air Force Base, New Mexico.

mately 140 gm. each by the start of the exposure period. All received two weeks' handling and gentling prior to the start of the experiment.

Apparatus

Exposure box.—This apparatus consisted of a rectangular, Masonite box with a milk-glass top. All sides and the bottom were painted white inside. A 60-w. bulb and an electric buzzer were placed above the glass top. Both of these were wired to a constant speed motor with a shaft and revolving disc to make possible the starting of the buzzer and light automatically at 30-sec. intervals for a 2-sec. period. Apparatus dimensions: 8 in. wide, 8 in. long, and 8 in. high.

Training box.—The training apparatus consisted of a hurdle-jumping device similar to that used by Kalish (5). It was 22 in. long, 6 in. wide, and 5 in. deep. There were two equal compartments provided with lids. The sections were divided by a 2-in. barrier. Both compartments contained a grid floor, a clear glass top, and were painted black inside. There was a buzzer and light attachment outside. The latency of each hurdling response was measured by a stop watch which was started when the CS was turned on and was stopped when the subject hit the floor of the opposite compartment.

Procedure

General.—The experimental design, arranged in three parts, was similar to that used by Bahrick (2). During the first part of the experiment, *Ss* were exposed to paired presentations of light and buzzer. In the second part they were given acquisition trials of the hurdle-jumping response with the buzzer as CS and shock as UCS to a criterion of 9 out of 10 avoidance responses. The third stage consisted of learning the same response to the light.

All *Ss* were under 22 hr. food deprivation during the SPC exposure period, but for the subsequent parts of the design the following experimental conditions were introduced. One experimental group (H) learned the avoidance response while under 22 hr. food deprivation. Another group (T) learned under 17 hr. water deprivation, and a third group (S) learned under satiated (hunger and thirst) conditions. The control group (C) learned the avoidance response under 22 hr. food deprivation. In measuring the transfer effect due to SPC, Group C's generalization was taken as an arbitrary zero level from which positive and negative amounts were computed for the experimental groups.

Exposure.—As noted above, all *Ss* were under the same deprivation schedule during this part of the experiment. The *Ss* were randomly assigned to one of the four groups to be used in this study. Each day for 14 days the three experimental groups were individually given 40 2-sec. exposures of light and bell every 30 sec. Group C was exposed similarly only to the test stimulus —the light. (Even though this precaution was properly carried out, Silver and Meyer's study [9] showed no differences in transfer among control groups given test stimulus or training stimulus or no stimulus in the exposure period.) Thus, the total number of exposures were 560 for all *Ss*. Seven exposure boxes were used simultaneously to facilitate running *Ss* in the exposure period. After each exposure session, *Ss* were kept in a delay box for 20 min. prior to being returned to their cages.

Training.—Each *S* was given hurdle-jumping training at the rate of 10 trials per day to a criterion of 9 out of 10 avoidance responses. The *Ss* had the buzzer as CS. The UCS (shock) followed the CS by 3.5 sec. and both were terminated automatically after the response was made or at the end of 60 sec. If *S* did not move on a given trial within 60 sec., he was removed and given a latency of 60 sec. for that trial. On any given day the order of running *Ss* was random, with the exception that Groups C and H were run prior to feeding and Groups T and S were run after feeding. Intertrial interval per *S* was maintained at a minimum of 10 min. to allow for the dissipation of reactive inhibition.

As stated above, one experimental group was under 22 hr. food deprivation, one under 17 hr. water deprivation,[3] and one was satiated. In order to insure against the possible generalization between hunger and thirst, the water-deprived *Ss* were given split feeding. They were fed for 1 hr. after 11 hr. of food deprivation and a second hour just prior to running. Thus at no time were they extremely hungry (the exposure condition) when they were extremely thirsty. Group C was also run hungry so that their results included the possibility for transfer learning based on the presence of autonomic processes similar to those in the exposure condition.

Test.—On the day following the end of the training series, all *Ss* were started on the test series with a running schedule similar to the above except for the following change. In this stage the light was used as the CS in place of the buzzer. The amount of generalization shown by Group C was used as the zero level for the purpose of making judgments of positive or negative transfer effects in the experimental groups.

[3] The 22 and 17 hr. food and water deprivation, respectively, were found to be approximately equivalent drive levels in an unpublished dissertation at the University of Pennsylvania, by O. W. Lacey, June, 1955.

TABLE 1

TRIALS TO NINE SUCCESSIVE AVOIDANCE CRITERIA ($N = 41$)

Group	Training		Transfer	
	Rank Sums	X^2_r	Rank Sums	X^2_r
C	27	5.53*	36	21.93**
T	20.5		25	
S	26.5		12	
H	16		17	

* $.10 < P < .20$, for 3 *df*.
** $P < .001$.

RESULTS

The principal basis for analysis of the transfer data was the number of trials to reach successive avoidance criteria. These data[4] for both training and transfer were subjected to analysis by Friedman's nonparametric analysis of variance (**8**). Although the groups did not differ in learning to avoid the buzzer ($.20 > P > .10$, Table 1), they transferred to the light differentially ($P < .001$). In Table 1, observation of the sums of the group ranks shows that the significance is principally due to the fact that Group C was consistently poorer in transfer than the experimental groups. (The reader will note that the sum of ranks in transfer for the control groups is the maximum possible when four groups are compared over nine criteria.) This result in turn is interpreted as evidence of a positive transfer effect for the experimental *Ss* using Group C as an arbitrary zero reference.

Further analysis by the Friedman technique shows that Groups T, S, and H differed in degree of positive transfer effect ($P = .006$). However, when the ranks of the groups were compared in pairs by the use of the binomial expansion test (**9**), only one comparison was significant. Group S (satiated) showed a consistently greater positive transfer effect ($P = .004$) than Group T (thirst).

Supplementary information regarding the most sensitive point in transfer at which one may expect the SPC effect is provided in Table 2. An analysis of variance by the Kruskal-Wallace method (**8**) shows that on the first 10 transfer trials the experimental groups gave more avoidance responses than Group C ($P < .02$). That this effect was not a result of initial free operant differences between these groups is indicated by identical analysis during training. The *Ss* did not differ on the first 10 training trials ($.95 > P > .90$). Applying the same analysis at the *end* of transfer showed that by the time the more stringent criterion of 9 out of 10 avoidance responses was reached, the above difference was no longer statistically significant ($H = 7.39$, $.05 < P < .10$).

Taken together, the data analyses reveal that the SPC effect appears almost immediately in the transfer stage, and, while it lasts throughout the rigid testing session (Table 1), it is strongest at the early stages of

TABLE 2

NUMBER OF AVOIDANCE RESPONSES IN FIRST 10 TRIALS ($N = 41$)

Group	Training		Transfer	
	Rank Sums	*H*	Rank Sums	*H*
C	219	.36*	313.5	9.87**
T	226.5		193.5	
S	193.5		176	
H	222		178	

* $.90 < P < .95$, for 3 *df*.
** $P < .02$, for 3 *df*.

[4] The original data are available in microfilm with University Microfilms, 313 North 1st Street, Ann Arbor, Michigan.

testing. Such an outcome is not difficult to understand if it is noted that with each successive transfer criterion the time since exposure increases. In fact, this occurrence suggests that for research in sensory preconditioning (SPC) massed testing will yield stronger effects than will distributed trials.

It is important to note again that the H, T, and S groups did not differ markedly in degree of positive transfer effect, and, if they differed at all, the satiated group tended to be superior. This means that the autonomic response complex present during exposure was apparently irrelevant and possibly even a hindrance in mediating the buzzer-light association at that time.

Discussion

If we attempt to interpret the present data in a mediation S-R framework, the most striking feature of the results is that varying the autonomic responses—presumably the major contributor to mediation in this instance—between exposure and the training-test stages had no detrimental effect on mediation. Indeed, the fact that Group S was superior to Group T suggests a tendency to the contrary. The very responses which should have facilitated mediation by being present during both exposure and the training and test stages actually seem to prevent mediation.

It would seem difficult in the face of the evidence presented to maintain a theory of learning such as Mowrer's (6) which is based solely on the mediation value of autonomic conditioning for subsequent skeletal learning. Because Osgood's S-R mediation hypothesis is more general, by virtue of a lack of restriction to any specific response, it is more difficult to anchor in any mediation study. Nevertheless, since Osgood cites many of Mowrer's experiments as classical examples of mediation learning, one may infer in the most conservative sense that, where they are consistent with stimulus association, autonomic responses may mediate learning. Consequently, in applying Osgood's mediation hypothesis to the design of the present experiment, one would be forced to predict some differences in transfer effect as a result of degree of autonomic similarity to that present during exposure. Again, if we ask what was the most consistent response complex present during the exposure period, we would be forced to state that it must have been either *solely* the autonomic activity associated with food deprivation or this activity in *combination* with some unobservable responses. Therefore, the most probable mediator from Osgood's mediated S-R viewpoint, as well as that of Mowrer, should have been the autonomic (and allied) hunger responses. The following analysis presents this interpretation in detail.

If we hypothesize, as do S-R theorists, that the PC stimuli are actually UCS's with unknown UCR's, then presumably prior to preconditioning the S-R relationships in the present study should have been as shown in Fig. 1. That is, the light and buzzer each elicited their own UCR and concomitant response-produced stimuli. Simultaneously, the induced autonomic (and allied) hunger responses (R_A) and their response-produced stimuli would also have been present. At the completion of Phase I, the preconditioning stage, a type of cross-conditioning between S_1 and S_2 would have been established with the result that each became a CS with a similar conditioned-response complex and response-produced stimulus complex. For example, the buzzer, S_1, now leads not only to r_1-s_1 but also to r_A-s_A and r_2-s_2. Now, in Phase II when S_1 is conditioned to R_3, so is the complex s_1-s_A-s_2. Consequently, in Phase III, the transfer test, the linking of S_2 to R_3 is mediated by its similar response-produced stimulus complex, s_2-s_A-s_1. If it is found, however, as it was in the present study, that r_A-s_A, the induced response complex, has no effect on mediation, the data would seem to pose some problems for S-R mediation theory. The difficulties are not obviated by an alternative interpretation given by S-R theorists which is simply to assert that the autonomic responses are not important for mediation.

In the latter case, there would be general agreement that Mowrer's mediation theory (6) does not hold. But if this is so, then it would seem somewhat damaging to Osgood's point of

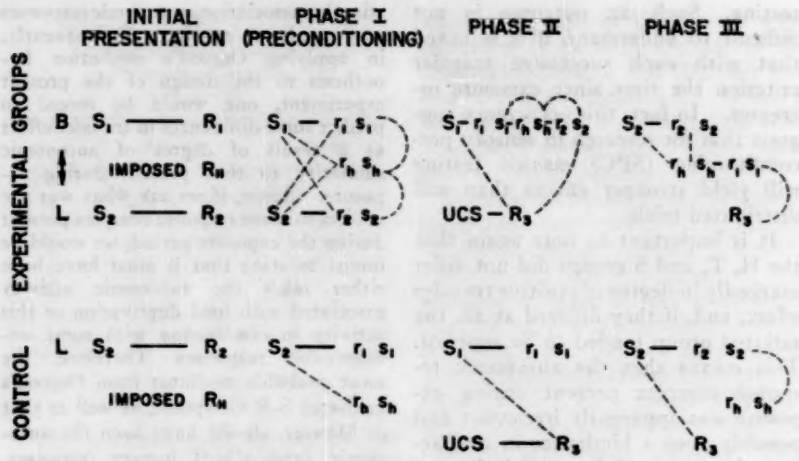

FIG. 1. A stimulus-response conditioning analysis of sensory preconditioning.

view (7) as well.[5] Although he does not restrict his theory to autonomic mediation, he has accepted Mowrer's data which seem to show that autonomic responses can be mediators in learning (7, pp. 398–412, 468–469). It seems inconsistent or at least theoretically not parsimonious to infer autonomic mediators in one transfer instance but to deny them to another where they are given equal opportunity to influence learning.

Pursuing this interpretation further, suppose it is granted that the s_h in the s_1-s_h-s_2 complex is not effective in mediation in the present study. The logical difficulty arises, then, of justifying the effectiveness of one inferred response-produced stimulus (s_1 or s_2) while discounting the value of another (s_h) which appears equally likely. In fact, in this experiment there was more certainty in the rat's induced hunger responses than in the existence of r_1 or r_2.

One other related point requires mention. During Phase I, it was assumed that a type of cross-stimulus conditioning was established between S_1 and S_2 and their unobserved UCR's. Since the parasympathetic autonomic responses, R_h, were also contiguous with the stimulus pairings, it is logical to conclude that the conditioned response complex after preconditioning consisted of r_2s_2-r_hs_h-r_2s_2. Note again Phase II in Fig. 1. When R_3 is conditioned to S_1, it is also conditioned to the response-produced stimulus complex associated with S_1, namely s_1-s_h-s_2. In the present experiment R_3 was avoidance behavior, a response complex which is dominated by the sympathetic nervous system. But the hunger responses in Phase I consisted primarily of antagonistic, parasympathetic reactions. In Phase II then, in order for the response-produced stimulus complex, s_1-s_h-s_2, to be associated with R_3, the antagonistic s_h associations with s_1-s_2 would have to be extinguished first. Such a process would require that the control group, lacking the antagonistic associations, learn the avoidance problem faster than the experimental groups. However, this did not happen.

While mediation S-R theory seems inadequate to account for the present data, S-S contiguity theory could predict such an outcome. According to Birch and Bitterman latent learning is simply a complication of SPC (4, p. 360). They thus could view the present study as a special case of the irrelevant drive paradigm. On the one hand, the hunger condition during exposure could be considered irrelevant since no appropriate reward is provided. Similarly, during training and testing the H, S, or T drive condition is extraneous to the avoidance task on hand. The necessary and sufficient conditions for mediation, S-S contiguity during exposure, were fulfilled for all experimental groups to the same degree (560 pairings of buzzer and light). With a sign relationship established between buzzer and light for the experimental groups, any subsequent instrumental response learned to one

[5] This interpretation holds equally for the "common response" mediation proposal (10). The only difference would be in the quality of the conditioned response complex; viz., it would be r_hs_h-r_es_e instead of r_1s_1-r_hs_h-r_2s_2.

PCS should be easier for them to transfer to the other PCS than for a control group with no pairing initially. Therefore, all should surpass the control group in transfer. The superior tendency shown by Group S could be rationalized from a cognitive viewpoint as being consonant with or analogous to the experiments on latent learning. The presence of a high irrelevant drive might simply prevent the animal from efficiently "attending" to the pertinent task.

From the foregoing analysis, the present data seem to provide an argument for the central interpretation of mediation in learning.

Summary

In the present experiment 41 hooded rats were run in a sensory preconditioning paradigm. After the preconditioning period, they were trained on an avoidance problem with one of the preconditioning stimuli as CS. The other stimulus was the CS in a transfer test. During the training-transfer stages the experimental *Ss* were divided into three groups based upon type of deprivation: hungry (H), thirsty (T), and satiated (S). The data showed a positive transfer effect for all experimental groups in relation to the control group. A slight superiority of Group S was indicated by its significantly higher positive transfer effect over Group T. It is concluded that in this SPC study specific responses were at least irrelevant and possibly detrimental to the mediation process. Further, the data are interpreted as favorable to an S-S contiguity theory of learning and problematical for an S-R mediation theory.

References

1. Bahrick, H. Latent learning as a function of the strength of unrewarded need states. *J. comp. physiol. Psychol.*, 1952, **45**, 192–197.
2. Bahrick, H. Sensory preconditioning under two degrees of deprivation. *J. comp. physiol. Psychol.*, 1953, **46**, 39–42.
3. Birch, H. G., & Bitterman, M. W. Reinforcement and learning: The process of sensory integration. *Psychol. Rev.*, 1949, **56**, 292–308.
4. Birch, H. G., & Bitterman, M. W. Sensory integration and cognitive theory. *Psychol. Rev.*, 1951, **58**, 355–361.
5. Kalish, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.
6. Mowrer, O. H. Two-factor learning theory: Summary and comment. *Psychol. Rev.*, 1951, **58**, 350–354.
7. Osgood, C. E. Method and theory in experimental psychology. New York: Oxford Univer. Press, 1953.
8. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
9. Silver, C. A., & Meyer, D. R. Temporal factors in sensory preconditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.
10. Wickens, D. D., & Briggs, G. E. Mediated stimulus generalization as a factor in sensory preconditioning. *J. exp. Psychol.* 1951, **42**, 197–200.

(Received September 9, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

A COMPARISON BETWEEN ROTE LEARNING AND CONCEPT FORMATION [1]

ROLLAND METZGER [2]

Northwestern University

Rote learning and concept formation have often been considered separate areas of investigation, but some writers, e.g., Underwood (**8**) and Osgood (**7**), put concept formation on the same dimension as learning, with conditioning near one end and thinking near the other. The present study will consider that portion of the presumed continuum where rote learning and concept formation overlap.

Many definitions of concept formation can be placed into either of two categories: (*a*) those in which stimuli grouped on the basis of common properties are attached to the same response, and (*b*) those in which two or more stimuli are associated with the same response but the stimuli themselves contain little or nothing in common. A relationship is then built up between stimuli by the fact that a common response is reinforced upon the occurrence of the stimuli. Concepts of both kinds will be examined in this experiment.

Osgood (**7**) indicates that the essential condition for concept formation is the learning of a common mediating response. A final common overt response can then be attached to the response-produced cue. In a diagram (**7**, p. 669) Osgood also points out that both discrimination and learning are required, first in the discrimination of the common characteristic among the stimuli and then in the learning of the nonsense syllable response to the common mediating response. In the following study both of the above processes of discrimination and learning may be involved in the concept-formation portion of the experiment. The *S* may discriminate a common element and associate a common response with it, or there is the possibility that *S* may in reality be using more of a rote learning component than a concept formation procedure in his learning, i.e., he may associate each individual response with each stimulus on a one-to-one basis without making use of any communality between stimuli. If the latter is the case, then there should be no difference between a rote learning group which learns one response to a stimulus and a group which learns the same response to two stimuli having common elements, provided responses do not have to be differentiated and number of response categories is not an effective variable.

Rote (paired-associate) learning is here defined as that type of learning in which *S* is required to associate a single response with a single stimulus, i.e., there is a one-to-one correspondence between stimulus and response. The term *concept formation* is here restrictively defined as that type of learning where *S* is required to associate the same response to different stimuli, i.e., two or more stimuli are attached to the same response. This restricted definition meets the char-

[1] This article is based on a dissertation submitted to the Graduate School, Northwestern University, in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his appreciation to his advisor, John W. Cotton, for guidance and encouragement during the course of the study.

[2] Now at Dixon State School.

acteristics implied in the two categories of concepts mentioned earlier.

Systematic concept formation indicates those conditions in which *E* attaches, to the same response, stimuli which have elements in common for the first two or three positions in the stimulus.

This study deals with several aspects of a stimulus-response learning task. Among these are number of elements in a stimulus, number of stimuli, number of response alternatives, concept formation, and rote learning.

In the design of this study a rote learning condition can be transformed into a concept formation condition in two ways: (*a*) by increasing the number of stimuli but keeping the number of response alternatives constant; and (*b*) by decreasing the number of response alternatives and keeping the number of stimuli constant. Experiment I will examine the first method; Exp. II, the second.

Experiment I

This experiment compares rote learning with systematic concept formation achieved by doubling the number of stimulus alternatives.

Method

Materials.—Individual stimuli consisted of groups of three or four triangles, each of which was either large or small. A typical pattern looked as follows: Δ Δ Δ. Responses were numbers ranging from 1 to 8 as needed. Each list consisted of 4, 8, or 16 stimulus-response pairs. Presentation was made by means of a memory drum. First the stimulus appeared alone for 2 sec. and then the stimulus and required response appeared together for 2 sec. (a 2:2 rate of presentation). Four random orders of presentation were used for each list. Learning was to a criterion of one perfect recitation.

Conditions.—There were four groups of 20 *S*s per group, each group being required to learn one of the lists shown in Table 1. In Table 1 the columns represent the type of learning variable (rote vs. systematic concept formation) and the rows the number of responses variable. The elements of the stimulus are represented in Table 1 by an "S" for a small triangle and an "L" for a large triangle. Thus Δ Δ Δ would be represented by LSL. As an illustration of what is meant by systematic concept formation, note in Group B that those stimuli which have LS as their first two elements both have the same response attached to them, viz., "1."

Temporal arrangements.—The *S*s were sequentially assigned to the groups, all with the

TABLE 1

Lists and Mean Trials to Reach One Perfect Recitation for Groups in Experiment I

Response Alternatives	Type of Learning		Row Means
	Rote	Systematic Concept Formation	
	A	B	
4	LSS-1 SLL-2 LLS-3 SSL-4 8.1	LSS-1 LSL-1 SLS-2 SLL-2 LLS-3 LLL-3 SSS-4 SSL-4 25.2	16.6
	C	D	
8	SLSS-1 LLSS-2 SSSL-3 SSLS-4 LLLS-5 LSSL-6 SLLL-7 LSLL-8 21.2	SLSS-1 SLSL-1 LLSS-2 LLSL-2 SSSS-3 SSSL-3 SSLS-4 SSLL-4 LLLS-5 LLLL-5 LSSS-6 LSSL-6 SLLS-7 SLLL-7 LSLS-8 LSLL-8 48.0	34.6
Column means	14.6	36.6	25.6

Note.—L indicates large triangle; S indicates small triangle.

understanding that they would serve for 1½ hr., even though it was apparent in some cases that this amount of time would not be necessary.

Results

The response measure used for this study was the number of trials required to reach the criterion of one perfect recitation. As some *S*s failed to learn the list to this criterion within 70 trials, a score of 70 was arbitrarily assigned to those *S*s. This truncation of the distribution, of course, makes for non-normality of the data, even after the transformation described below. Cochran (3) has pointed out that no serious error is introduced by non-normality in the significance levels of the *F* test.

Table 1 shows the mean number of trials to reach the criterion for each of the four groups in this experiment.

A Bartlett's test (3) for homogeneity of variance was significant ($\chi^2 = 11.36$ for 3 *df*). The transformation $X' = \sqrt[4]{X}$ was therefore applied to each *S*'s score. A Bartlett's test on the transformed data resulted in a nonsignificant χ^2 (2.47 for 3 *df*), permitting the conclusion that the X′ distributions have homogeneous variances. The transformed data were used in all further analyses.

The results of an analysis of variance on transformed scores for the groups of Table 1 show significant row effects ($F = 63.10$; 1 and 76 *df*), significant column effects ($F = 94.98$; 1 and 76 *df*), and no significant interaction ($F < 1$; 1 and 76 *df*). This analysis of variance indicates that the effect of rote learning vs. systematic concept formation is significant, when the difference between these two types of learning is achieved by doubling the number of stimuli while keeping the number of response alternatives constant.

Experiment II

In contrast to Exp. I, this experiment compares rote learning with systematic concept formation achieved by altering the number of responses while the number of stimuli are held constant.

In addition, two other comparisons will be made, viz., the effect of an additional column element to the stimulus and a comparison between systematic concept formation and random concept formation. The term *random concept formation* indicates those instances in which the stimuli attached to the same response by *E* are assigned at random. This is in contrast to systematic concept formation in which stimuli attached to the same response by *E* have the same elements for the first two or three positions.

Method

The materials were identical to those in Exp. I. There were eight groups of 20 *S*s per group, each group being required to learn one of the lists shown in Table 2. Groups B and C were already reported in Exp. I. The temporal arrangements were identical to those in Exp. I.

Results

The response measure and maximum possible score for Exp. II are identical to those of Exp. I. Table 2 shows mean number of trials to reach the criterion for each of the eight groups of this experiment.

To make all comparisons in this experiment on the same basis, a Bartlett's test was made for all eight groups of Table 2 and it was found that the sample variances differed significantly ($\chi^2 = 30.91$ for 7 *df*). The transformation $X' = \sqrt[4]{X}$ was therefore applied to each *S*'s score. A Bartlett's test on the transformed data resulted in a nonsignificant χ^2

TABLE 2

LISTS AND MEAN TRIALS TO REACH ONE PERFECT RECITATION FOR GROUPS IN EXPERIMENT II

Response Alternatives	3 Elements 8 Stimuli	4 Elements 8 Stimuli	4 Elements 16 Stimuli
	E	F	G
4-Randomly assigned	LLL-1	LLLS-1	SSLL-1 SLSL-3
	LSS-1	LSSL-1	SLLS-1 LSSS-3
	SSS-2	SSSL-2	SLLL-1 LSLS-3
	LSL-2	LSLL-2	LSLL-1 LLLS-3
	SLS-3	SLSS-3	SSSL-2 SSSS-4
	LLS-3	LLSS-3	SSLS-2 SLSS-4
	SLL-4	SLLL-4	LLSL-2 LSSL-4
	SSL-4	SSLS-4	LLLL-2 LLSS-4
	18.4	40.4	51.2
	B	H	J
4-Systematically assigned	LSS-1	LSSL-1	LSSS-1 LLSS-3
	LSL-1	LSLL-1	LSSL-1 LLSL-3
	SLS-2	SLSS-2	LSLS-1 LLLS-3
	SLL-2	SLLL-2	LSLL-1 LLLL-3
	LLS-3	LLSS-3	SLSS-2 SSSS-4
	LLL-3	LLLS-3	SLSL-2 SSSL-4
	SSS-4	SSSL-4	SLLS-2 SSLS-4
	SSL-4	SSLS-4	SLLL-2 SSLL-4
	25.2	25.0	38.4
	K	C	
8	SLS-1 LLS-2 SSS-3 SSL-4 LLL-5 LSS-6 SLL-7 LSL-8	SLSS-1 LLSS-2 SSSL-3 SSLS-4 LLLS-5 LSSL-6 SLLL-7 LSLL-8	
	19.0	21.2	

Note.—L indicates large triangle; S indicates small triangle.

(6.96 for 7 *df*). All further analyses were made on the transformed data.

An analysis of variance was made on the transformed data of the 2 × 2 table formed by Groups B, H, K, and C, in which the rows represent the systematic concept formation vs. rote learning variable and the columns represent number of elements in the stimulus. This analysis indicates no significant differences for type of learning ($F = 3.85$; 1 and 76 *df*), for number of elements in the stimulus ($F < 1$; 1 and 76 *df*), or for interaction ($F < 1$; 1 and 76 *df*). It may be concluded that when a rote learning task is changed to a systematic concept formation task by reducing the number of different response alternatives, the rate of learning is not changed significantly.

The comparison between systematic and random concept formation was made as two subanalyses. The first has the number of elements in the stimulus as the second factorial variable; the second subanalysis has the number of stimuli as the second factorial variable.

The first subanalysis consists of the 2 × 2 table formed by Groups E, F, B, and H where rows represent the type of concept formation variable and columns represent number of elements in the stimulus. This analysis of variance shows the main effect of type of concept formation not to be significant ($F < 1$; 1 and 76 *df*) while number of elements in the stimulus ($F = 10.22$; 1 and 76 *df*) and interaction ($F = 13.40$; 1 and 76 *df*) are significant.

The second subanalysis consists of the 2 × 2 table formed by Groups F, G, H, and J where rows again represent the type of concept formation or number of responses variable but the columns now represent number of stimuli. An analysis of variance

shows both type of concept formation ($F = 11.63$; 1 and 76 *df*) and number of elements in the stimulus ($F = 11.52$; 1 and 76 *df*) to be significant, but interaction not to be significant ($F < 1$; 1 and 76 *df*).

Discussion

The main finding of Exp. I is that, when a rote learning task is transformed into a systematic concept formation task by doubling the number of stimuli and keeping the number of response categories constant, the difficulty increases. This statement should be critically examined in the light of the operations involved. Doubling the number of stimuli naturally doubles the length of the task, since the length of the task is intrinsically associated with the number of stimuli. Therefore, when we are manipulating rote learning versus systematic concept formation by altering number of stimuli, we are almost by definition varying length of list. Length of list has long been known to be an effective variable (2, 5, 6) with longer tasks being more difficult. This conclusion, then, is a special case of a well known principle.

The significant increase in trials to learn with an increase in number of responses (the row variable of Table 1) must not be interpreted as a pure effect of number of responses. It will be noted in Table 1 that the rows differ not only in number of responses but also in number of stimuli and number of elements in the stimulus. Therefore this row effect really represents a combined effect of number of stimuli, number of elements, and possibly number of responses.

One of the purposes of Exp. II was to determine whether there was any difference between rote learning and concept formation when concept formation was achieved by reducing the number of different response categories. The analysis of variance of Groups B, H, K, and C indicate that a change in number of response categories is not an effective variable.

The result of this analysis is in line with that of Bricker (1), who used binary stimulus elements and changed the number of responses available to *S*. He found that his perceptual task did not seem to be affected by changes in number of available responses. When the number of stimuli was held constant, in the present experiment and in Bricker's experiment, variation in number of response categories was not found to be a significant variable.

The result that number of response alternatives is not an effective variable also agrees with an experiment by Hake and Eriksen (4). Hake and Eriksen had *S* attach varying numbers (2, 4, or 6) of labels (responses) to a series of 16 visual patterns. These writers found differences in learning as a function of number of responses in their task, which was somewhat similar to that of the present study; the difference, however, disappeared when their raw data were corrected for chance responses.[3]

It could possibly be expected that number of response alternatives in a task should be an effective variable when the responses have to be acquired for the first time as, for example, in a nonsense syllable task. Note that the responses for the current study were so chosen that *S* would *not* have to learn which response options were available. The *S* already knew the responses (numbers) per se extremely well. The *S*'s task was to *associate* the response with the proper stimuli. Therefore this study tested the association of stimulus and response unconfounded with the acquisition of a set of responses. Under these conditions there was no difference between rote learning and systematic concept forma-

[3] No corrections were made for chance responses in the present experiment due to the type of response measure used, i.e., trials to criterion. In a four-item list where chance responses would be at their highest effectiveness, the probability of *S* attaining the criterion of four consecutive correct responses by chance alone would be $(1/2)^4 = .0625$. This was considered to be so low as not to warrant correcting for chance responses. In an 8- or 16-item list, of course, the probability would be even lower.

tion when the number of response categories was varied.

A set of postulates which might account for the results of Table 2 are as follows:

1. In Groups E, B, and K (three column) the task is relatively simple so that factors which might make for difficulty have no effect and no use is made of any help which might be available to overcome any factors which might make for difficulty in a more complex task. Hence Groups E, B, and K tend to be relatively alike.
2. In a more difficult task (four column stimulus) response interference (two or more stimuli attached to the same response) tends to increase difficulty.
3. Systematic concept formation tends to reduce difficulty.
4. An increase in number of stimuli increases difficulty.

Postulate 2 accounts for the high value of Group F. Postulates 2 and 3 together account for the relatively low value of Group H. The difficulty created by response interference is offset by the aid afforded by systematic concept formation. Group C has a relatively low value because there is no response interference. Postulate 4 accounts for the higher values of groups G and J as compared with F and H, respectively.

With the above postulates, the nonsignificant difference between rote learning and systematic concept formation in Exp. II would be interpreted in the case of the four-element stimulus groups as the net effect of response interference offset by the help afforded by systematic concept formation. The fact that systematic concept formation has an effect is shown by the significance of the variable of type of concept formation in the second subanalysis above.

In summary, it may be stated that the main result of Exp. II is that when a rote learning task is changed to a concept formation task by varying the number of response alternatives, the change in number of response categories is not a significant variable.

Summary

This investigation compared rote (paired-associate) learning, systematic concept formation, and random concept formation.

Experiment I compared rote learning with systematic concept formation when the difference between the two types of learning was achieved by varying the number of stimuli. The difference was found to be significant. The change in number of stimuli was identified with length of list which has long been known to be an effective variable.

Experiment II compared rote learning with systematic concept formation when the difference was achieved by varying number of responses. This difference was not found to be significant. Experiment II also compared systematic concept formation with random concept formation in two factorial designs. No difference was found between these two types of concept formation when number of elements in the stimulus was the second factorial variable. The difference between the two types of concept formation was, however, found to be significant when number of stimuli was the second factorial variable.

References

1. Bricker, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.
2. Ebbinghaus, H. *Memory: A contribution to experimental psychology.* Trans. by H. A. Ruger and C. E. Bussenius. New York: Bureau of Publ., Teachers Coll., Columbia Univer., 1913.
3. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Hake, H. W., & Eriksen, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.
5. Hovland, C. I. Experimental studies in rote-learning theory. VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27**, 271–284.
6. Meumann, E. *The psychology of learning.* Trans. by J. W. Baird from 3rd German ed. New York: Appleton, 1913.
7. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
8. Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

(Received September 30, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE INTERACTION OF PITCH AND LOUDNESS DISCRIMINATIONS

J. DONALD HARRIS, ANDREW G. PIKLER, HOWARD S. HOFFMAN,[1] AND RICHARD H. EHMER

U. S. Naval Medical Research Laboratory

In a previous study from this laboratory, Hoffman (3) showed that a noise was equally detectable no matter whether it was steady, interrupted, or frequency modulated. However, the detection of these attributes in the signal was affected by certain combinations; it was shown that the detection of the presence of periodic interruptions in a noise was unchanged when frequency modulation was added, but that the detection of the presence of frequency modulation was more efficient without periodic interruptions than when these were introduced. It would seem that combining two different attributes in a signal can render one of the attributes even less apparent. In Hoffman's experiment (3) the periodic interruptions may have broadened the bandwidth of modulation, but some sort of psychological effect also may have occurred. One must also be prepared to expect enhanced effects as well as reduced effects under some conditions of attributes to be added.

A number of interesting questions in cue detection and signal discrimination are raised by this type of approach. How can an attribute, with a certain discrimination-probability for a simple change in a steady state, *not* add its bit to the discrimination-probability for a complex change? Is there any real inhibitory process involved? Should one not oftener expect a facilitation? What is the size and nature of the correlation between the psychophysiological processes underlying two attributes presented simultaneously? What is the expected statistical coupling between two uncorrelated attributes?

It was decided to attack these problems with somewhat simpler stimulus complexes and simpler response categories than used in the previous pioneer study. A steady-state pure tone was varied (after a brief silent interval) either in frequency or intensity, or in both simultaneously. The *S* was not required, as previously, to decide whether either or both attributes varied, but simply to discriminate differences in tones which varied slightly in frequency and/or intensity.

The problem thus becomes one of determining the joint discrimination-probability of the combination of two and only two simultaneous cues, the discrimination-probabilities of each cue alone having been rather precisely determined beforehand.

Pitch and loudness cues were chosen utilizing an interval of silence, rather than any of dozens of other possibilities, because a series of factor analytic studies (2) have provided a tentative notion of the size of the correlation between these two primary auditory abilities, among others. This correlation appeared to be useful in helping to interpret discrimination-probabilities yielded by the two cues simultaneously.

The generalized problem, in somewhat different form, has been dealt with before. For two cues, or two sensory inputs which are uncorrelated, one would necessarily expect that what is

[1] Now at the Pennsylvania State University.

missed by one cue, or sensory input, will be picked up with a certain detection-probability by the other cue or input. Barany (1) and Pirenne (6) considered this possibility for the two sensory inputs represented by binocular vision. Smith and Licklider (7) considered contact detection in quiet with one vs. two ears. Pikler and Harris (5) extended the binaural problem from contact detection to pitch discrimination, and wrote the formula for the joint discrimination-probability of two simultaneous uncorrelated cues:

$$P_{joint} = P_1 + P_2 - P_1P_2 \qquad [1]$$

This formula will be referred to as the statistical coupling hypothesis. In the last-mentioned study the authors leaned toward rejecting the statistical coupling hypothesis in favor of a dynamic interaction between the ears.

Marill (4) has considered the joint detection-probability of two frequencies separated by a relatively slight frequency difference (40 cps) and again separated by a full octave. He shows that for the octave separation the joint probability is simply that of the more detectable frequency, but that for the narrow separation the less detectable frequency yields a significant increase (maximum improvement: 82% and 84% add up to a joint 96%; minimum improvement: 61% and 62% add up to a joint 65%). This increase varies with percentage difference between frequencies.

This means that for the human observer it matters whether a signal energy is spread over two (or more) components, or is contained within a single pure tone. (As Tanner [8] has shown, the human observer acts something like a subjectively adjusted filter in that the detection-probability of a frequency is less when that frequency is only one of two frequencies which may be presented—and this selective filter action is sharp enough to register a 10% frequency difference.)

Marill showed that the detection of a two-component signal, both components of which are within the pass-band of the human filter, can be well predicted by considering the total energy to be concentrated in either one of the two components; one simply combines energies, enters the graph for either frequency to find the detection-probability at that energy, and the datum is a good prediction of the joint detection-probability.

It is not, unfortunately, possible to use this relatively simple technique in predicting joint discrimination-probability when two immiscible cues or attributes of a signal, such as frequency and intensity, are simultaneously altered.

The present experiment was devised to obtain the joint discrimination-probabilities of a large number of combined pitch and loudness cues, to relate the joint probability of each combination to those of the cues individually, and to explore so far as possible the nature of the interactions, if any, of these cues.

A relatively low frequency was selected in order to work within a frequency region where the "place" theory of pitch has least cogency. At 125 cps it is thought that frequency as well as intensity is mediated in some way by number of synchronously active fibers. Presumably the correlation between attributes should be highest where similar neural mechanisms are involved.

Method

Five adult men with normal acuity at 125 cps were used. They were required to make 120 judgments at a sitting as to whether a third tone was more like the first or the second of two previous tones. This is the now familiar "ABX" technique. Tone A was always a 125-cps tone 1 sec. in duration at 10 db sensation level. Tone B, following 1 sec. of silence, itself also 1 sec. in duration, was either 0, .5, .75, 1, or 1.25 cps greater in frequency, as well as 0, .4, .8, 1.2, or 1.6 db higher in intensity. Each of these 24 combinations (0–0 was never given) was presented five times in each session in a random fashion. After another 1 sec. of silence, Tone X was presented, which was identical with either A or B in a random pattern. A 5-sec. voting period followed. This sequence was impressed twice on magnetic tape, Forms A and B, with a good Ampex studio tape recorder, using the usual precise laboratory switches, attenuators, etc., and was played back to *S* with the same Ampex equipment.

The range of cues was carefully chosen by previous exploration so that each pitch and loudness cue, by itself, fell within each *S*'s interval of uncertainty, and preferably nearer to 50% than 100%. Two compromises were made: in the least discriminating *S* the cues were made generally easier on another tape, and in the most acute *S* the sensation level was dropped to 7 db. All *S*s generally made 100 judgments on each combination. One *S* needed only 50 judgments,

and one S needed 200, before an informal sequential analysis revealed anything like satisfactory stability.

Results

The raw data, in the form of percentage of correct items for each combination, are entered in Table 1.[2] The cells represent each a certain combination of pitch and loudness cues. The cells in the upper left quadrant represent those combinations relatively easy in both dimensions; the upper right quadrant contains combinations of easy pitch cues and relatively difficult loudness cues; and so on. Note that the extreme right hand vertical column represents pitch cues only (the loudness cue being 0), while the bottom row represents loudness cues only (pitch cue of 0).

The matrix of cells allows one to look at the discrimination-probability of, say, a 1-cps frequency change and to observe the progressive effect of adding a more apparent loudness cue (read from right to left, next to top row). All other progressions can similarly be brought to light by passing through the matrix using other pathways.

Judgmental difficulties.—The S experiences many difficulties in discriminations of the sort demanded here. In the first place, the ABX judgment demands, or at least makes possible, a multiplicity of subjudgments before a final judgment is made. However, the crucial advantage of the ABX technique in this particular experiment (preventing S from responding "Same" in the usual AX technique) far overrides the disadvantage of judgmental complexity.

The optimum situation, it might be thought, would be perhaps to give S two tries at every item—once when he attends to the pitch cues and once to the loudness cues. In the present case this could not be done for two reasons: (*a*) with a great many items S was unaware of any change at all (though he could still perform better than chance), and (*b*) in many cases a slight change in intensity seems to

TABLE 1

Percentage Correct Judgments for Combinations of Pitch and Loudness Cues

Pitch Cues in Descending Order of Difficulty	Variable Differences in Decibels				
	1.6	1.2	.8	.4	0
4	S_1: 82	87	85	77	78
	S_2: 95	92	94	93	88
	S_3: 97	74	86	72	59
	S_4: 74	71	72	57	69
	S_5: 91	90	94	80	78
3	81	79	76	84	70
	95	83	88	83	72
	85	82	88	67	58
	66	67	75	63	63
	94	96	75	80	76
2	84	74	80	72	64
	88	89	87	84	85
	94	86	70	72	57
	63	66	63	69	53
	86	90	87	68	70
1	77	83	67	67	62
	89	91	79	76	79
	84	76	70	57	59
	62	67	54	62	55
	98	82	80	68	46
0	65	61	65	43	
	68	71	66	43	
	85	69	63	59	
	62	62	55	57	
	95	76	74	60	

Note.—For S_5 the pitch steps were 1.5, 1.75, 2.0, and 2.25 cps. For all others the steps were .5, .75, 1.0, and 1.25 cps.

[2] Tables giving for each S (*a*) the obtained percentage correct, (*b*) the percentage predicted from Formula 1, (*c*) the datum smoothed in the pitch dimension, (*d*) the datum smoothed in the loudness dimension, and (*e*) the average smoothed data, have been deposited with the American Documentation Institute. Order Document 5640, remitting in advance $1.25 for microfilm or $1.25 for photocopies.

produce a pitch change, and vice versa, so that the attributes are not in any real sense separable.

Another difficulty experienced by all *Ss* was concentrating for long periods on the very weak sounds used, and on the minute differences introduced (the level of certainty was by design kept extremely low to reduce the expected ceiling effect, if any additive effect should appear).

In view of these difficulties it was not surprising that there should be some inversions in the data among adjacent cells. Adjacent cells in this study, after all, represent to *S* almost infinitesimal differences in any dimension.

It was clear from the analysis of the data when half completed that additional judgments were likely to aid statistical clarity, and a complete replication was performed. An item analysis, and a comparison of data from Forms A and B, eliminated two items which were either miscut or subject to some obscure serial position effect. It was felt at this point that additional judgments would not add significantly to our knowledge of the nature of cue interaction, and the experiment was terminated.

Reliability of the data.—There are two ways, at least, to look at the reliability of the data. One can consider the judgments within each cell as successive series. This has been done for successive groups of 20 judgments each, calculating the *SD* of the mean for each cell using a small-samples formula. The *SD*'s for the 0 rows and columns run about 10% (*mdn.* = 11.2%) and are representative of the whole matrix. It does not appear that the variance of cells containing one cue alone is any different from that of cells containing two cues.

It is clear that differences between any two adjacent cells are statistically unclear, as they are phenomenologically unclear (as may be expected when one is working within such a small portion of the interval of uncertainty), but that for larger and larger differences along any dimension a higher and higher level of confidence is achieved.

Another look at reliability in a general sense is provided by the consistency of the two "smoothing" processes available for every cell, horizontally and vertically. If each row and column is considered as a set of psychophysical data, a curve can be fitted to each row and column, which yields two "smoothed" figures for each cell. These values (data fitted by probit paper) are for every *S* for every cell. Now in a cell where the two coincide, and agree with the raw data, one has, of course, the most confidence in the reliability of the raw data. Where they coincide, but the raw data diverge in one of the cells, one tends to feel that the raw data in that cell were subject to fluctuation from unknown causes. Where the two smoothed figures disagree, and the raw data differ from either, one seeks to add additional raw data. In 119 of the 144 cells the raw data agree within 5% or less as compared with the average of the two smoothed frequencies.

Generalization of the usual two-variable psychometric function.—By adding a second independent variable, the plane geometry of the psychometric function is extended to a third dimension. The variable tone is presented as different both in frequency and intensity within the two separate intervals of uncertainty. The new dimension leads to the concept of "psychometric surfaces" rising above the plane of the usual two-dimensional plot. In this situation as in the

classical two-dimension situation the dependent variable is percentage of correct answers.

A first clear observation is the generally ordered nature of all matrices. For all *S*s there are generally consistent progressions in discrimination-probabilities as one moves in Table 1 from the bottom upwards, or moves from right to left. This indicates (*a*) that some sort of interaction (defined later more carefully) does occur, and (*b*) in an additive direction (since the joint probabilities are almost always more than those of the more discriminable cue), and (*c*) that this interaction is a general characteristic for all *S*s.

A second observation is that in almost no cell does the experimental datum equal or approach the figure predicted by the statistical coupling of uncorrelated cues. This comes as no surprise, since it was already known that the cues used here were not uncorrelated (**2**). The joint probability of any cell can be calculated from Formula 1. (These figures may be described as "probabilities" because they can be assumed to be related in a linear fashion to the true probabilities. With the present data, 50% correct is chance and has a probability of 0; 75% correct has a probability of .5; and so on.)

A third observation is that there is no consistent tendency for a cue to alter its percentage contribution when the value of the other component changes. If, for example, a pitch cue of .5 cps yields a percentage of 10 better than chance, then one can expect that adding a pitch cue of .5 cps to *any* loudness cue (limits discussed below) will result in an improvement of just about 10% in the joint probability over that for the loudness alone. This is, of course, true only for the relatively linear portion of the psychometric curve for loudness alone.

With S_1 for instance, for whom the joint probabilities all are encompassed between 67% and 87%, and with S_4, for whom the joint probabilities are also quite low, the percentage over chance furnished by a second cue alone is a uniformly good predictor of the percentage increase furnished by that second cue to the joint probability. For S_2, for whom the joint probabilities are high, ranging from 76–95%, and generally are over 85%, the percentage contribution of a second cue is truncated by a ceiling effect. But even in S_2, the full percentage contribution increase expected from the second cue is found wherever the first cue is relatively weak.

A seemingly contradictory *S* in this respect is S_3, in whom not less, but *more* is occasionally contributed by a second cue than would be expected. But note that S_3 is operating generally in the initial portions of his psychophysical functions; in these portions a certain stimulus increment occasions a *slighter* rise in probability of discrimination than in later portions.

The possibility here is helpful of equating the effects of frequency and intensity one to the other by going through the psychophysical functions as transforms.

S_1 is an illustrative case in which the truly linear portions of the psychophysical functions are available. For S_1 an increase of 10% in the middle portion of his pitch curve is occasioned by a pitch cue of about .55 cps, and by a loudness cue of about .62 db. Thus, in terms of a unit on the psychophysical function, an increment of .55 cps is equivalent to one of .62 db. Curve A (Fig. 1) can now be adjusted by .55 cps instead of 10%, as is done in Curve B. Curve B, now,

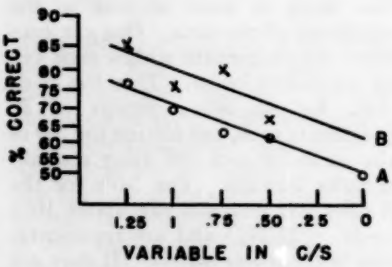

FIG. 1. The obtained psychophysical function for pitch (circles) and the best fitting ogive (Line A); the obtained effect of adding .8 db intensity cue to pitch cues (crosses) and the predicted effect of adding .62 db intensity cue to pitch cues (Line B).

represents a prediction of the addition of .62 db loudness cue to all pitch variables. The crosses entered in Fig. 1, which represent the raw probabilities of pitch-cues-plus-.8-db-loudness-cue (as close as available here to .62 db) are seen to follow the prediction within experimental error.

Discussion

From the above it is known that interaction occurs and always in an additive direction, and almost never reaches the level predicted by the hypothesis of statistical coupling with uncorrelated cues. These points will be discussed in order. By the noncommittal term "interaction" we mean only that the two simultaneous cues produce results different from either cue alone, and different from the results predicted by the statistical hypothesis. If the compound cue yielded only the datum of the more powerful cue, we would have to conclude that the less powerful cue was completely disregarded; whereas if the compound cue yielded the datum predicted by the statistical hypothesis, we would have to conclude that the two cues were completely uncorrelated. Since the truth quite evidently lies between these extremes, we conclude that some interaction exists.

We offer two possible explanations for this interaction, or rather we conceive two synergistic interactions, one in terms of a synchrony or partial synchrony between fluctuations of discriminability for pitch and fluctuations of discriminability for loudness, the other in terms of an overlap of cochleo-neural events for pitch and for loudness. We believe that both interactions occur, and we provisionally call the first type "correlation of fluctuation phase," and the second type "cochleo-neural overlap."

A positive correlation of fluctuation phase would occur if peak sensitivity for pitch and peak sensitivity for loudness occurred simultaneously, and minimum sensitivity likewise. If the two fluctuations were at random one with the other, no correlation would occur. These two fluctuations undoubtedly do occur, as we see from the uncertainty of the psychophysical function, but one can hardly believe that they are completely uncorrelated, since they have a convergence in the central nervous system.

It might be thought that this type of correlation alone could account for the data, except that this type of interaction would have the same magnitude no matter what the individual probabilities of the cues in any combination; in other words, the amount by which the raw data would fall below the statistical prediction should be a constant. Yet Table 1 shows that the difference between the obtained and predicted probabilities varies systematically through the matrices. The differences are least in the upper left-hand quadrants and greatest in the lower right-hand; and this is true even for S_1 and S_4, for whom no ceiling effect is present. We conclude that at least one type of interaction must be in play in addition to the correlation of fluctuation phase, and moreover is dependent in a regular manner upon the strength of the cues, being greatest when the strength is weak.

We here suggest that the second type of interaction is dependent upon cochleo-neural overlap. When a pure tone changes in the region of 125 cps, even by only a fraction of a cycle, a slight shift in the cochlear pattern basalwards may be expected on the basis of electrical phase studies of cochlear patterns. But this is also exactly what one expects in the case of a slight increase in intensity, as is known from the basalward asymmetry of masking data especially at low frequencies. We believe that this type of interaction is a consequence of the electroanatomy of cochlear patterns and that for these two particular cues only an additive interaction could occur.

In order to explain these results, the present theory is that the two types of interaction operate in the same additive direction, the correlation of fluctuation phase having a low but constant value, the cochleo-neural overlap having a variable value depending upon strength of individual cue.[3]

[3] The second author does not entirely agree with the foregoing discussion, preferring to re-

A suggested formula to contain these ideas is as follows:

$$P_{joint} = P_1 + P_2 - kP_1P_2 - C \quad [2]$$

where C is a constant depending upon the correlation of fluctuation phase, and k is a constant depending upon extent of cochleo-neural overlap. From the present data, it is possible to estimate that C is of the order of magnitude of 5% (about the minimum reduction of obtained from predicted probabilities for M and P, the two *S*s with little or no ceiling effect), and k is of the order of 1.2 for M and 1.4 for P (there is the strong possibility of a real *S* difference in extent of interaction).

One can conceive further studies which will throw light upon these matters. Until some of the parameters are explored, the present explanation of the nature of interaction must be regarded as speculative.

Summary

The interaction of pitch discrimination and loudness discrimination was studied when both cues were present in the same stimulus.

The addition of a second cue, even when it was by itself almost completely undetectable, never failed to add its bit to the joint discrimination-probability of a two-cue signal. This interaction effect was fairly constant no matter what the discrimination-probability of the first cue.

The interaction effect almost never reached the level expected on the hypothesis that what items are missed by one cue will be recovered with a certain probability by the second cue. It was concluded that the amount by which an obtained discrimination-probability fell short of the prediction, was a measure of the interaction or interactions between the two cues involved. The relative size of the interaction seemed highest when the two cues were weak (i.e., had individually low discrimination-probabilities).

It was speculated that the nature of this interaction was twofold, and was to be looked for in a correlation between sensitivity fluctuations for pitch and loudness, and especially in the anatomico-electro-neural events within the cochlea which move basalward with increased intensity and also with higher frequency.

gard the task as one of problem-solving involving decision-making on a statistical basis.

References

1. Bárány, E. Some statistical observations on the methods in threshold determinations in general with particular regard to determination of visual acuity and subliminal addition. *Acta Ophthal., Kbh.*, 1946, **24**, 123–127.
2. Harris, J. D. A search toward the primary auditory abilities. *U. S. N. med. Res. Lab. Rep. 57-4.* April, 1957.
3. Hoffman, H. S. The detection of signals and their attributes. *U. S. N. med. Res. Lab. Rep. No.* 277. Sept., 1956.
4. Marill, T. Detection theory and psychophysics. *M. I. T. Res. Lab. Electronics Tech. Rep. No.* 319. Oct., 1956.
5. Pikler, A. G., & Harris, J. D. Channels of reception in pitch discrimination. *J. acoust. Soc. Amer.*, 1955, **27**, 124–131.
6. Pirenne, M. H. Binocular and uniocular threshold of vision. *Nature, Lond.*, 1943, **152**, 698–699.
7. Smith, M. H., & Licklider, J. C. R. Statistical bias in comparisons of monaural and binaural thresholds: Binaural summation and supplementation. *Psychol. Bull.*, 1949, **46**, 278–284.
8. Tanner, W. P., Jr. Theory of recognition. *J. acoust. Soc. Amer.*, 1956, **28**, 882–888.

(Received October 16, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE EFFECTS OF RESPONSE CONTINGENT PROBABILITIES WHICH FAVOR RESPONSE CHANGE [1]

MAYNARD W. SHELLY, II [2]

University of Wisconsin

There has been an increasing number of studies concerned with responses made to the relative frequencies of events. A general hypothesis to come out of these studies, when *S*s can make one of two responses, is that they tend to asymptotically match the relative frequency of reinforcement with a relative frequency of responding (**3, 6**). Several exceptions to this hypothesis have already appeared in the literature. Among these are studies which have employed more than two response alternatives (**1, 5**).

The present study extends the investigations of responses to relative frequencies. It explores the response to relative frequencies in situations where these relative frequencies depend upon the preceding response. Such relative conditional frequencies may be interpreted as response contingent probabilities.

The primary purpose of this study is to investigate certain conditions which favor response change on successive trials. The possible importance of this variable in defining restrictions upon the probability matching hypothesis results from the type of disagreement usually found. Most disagreements with this hypothesis are in the direction of a decreased probability of changing responses on successive trials. With the exception of the control condition, all probabilities used in this experiment led to a reinforcement only when *S*'s response differed from his preceding response.

A second purpose is to investigate the extent to which conditional relative frequencies are matched by response probabilities. The situation used presents *S* with four possible alternative responses.

This experiment has, additionally, another interpretation. Situations in which the relative frequencies of events are dependent upon preceding responses might be considered to be an elementary prototype of social situations. One of the most characteristic features of social behavior is that the consequences of a person's behavior depend partly upon preceding events which often consist largely of his own responses. Since in this experiment an apparatus has been constructed which changes the probabilities of an individual's success as a function of his preceding response, the situation investigated may be considered to be an elementary social situation.

METHOD

Task.—The *S*s in this experiment were told that they were playing a game against a machine and that on each trial either they or the machine

[1] This article is based on data contained in a dissertation submitted to the faculty of the Department of Psychology of the University of Wisconsin in partial fulfillment of the requirements for the Ph.D degree, June, 1956. The author wishes to express his appreciation to the Social Science Research Council for a fellowship to carry out this research as part of a research program, and to J. C. Gilchrist for his stimulating ideas concerning the importance of response conditional probabilities and for his constant encouragement, suggestions, advice, and patience.

[2] Now at the Laboratory of Aviation Psychology, Ohio State University.

would win. The object of the game was for them to win as often as possible.

Apparatus.—The apparatus was part of a more general apparatus designed for studying various forms of stimulus-response association learning and probability learning. The *S* sat in a booth which was painted flat black. The booth contained an aluminum chassis box (8 × 17 in.) which was mounted on a sloping table. This box contained 16 four-pole snap switches, above each of which was mounted a 6-v. green-domed lamp. For the present study, only the four switches at the lower right of the box were used. The remainder of the switches and their associated lamps were covered in a way that permitted free access to those switches not covered. Another aluminum chassis box was mounted on the wall in front of *S*. This contained an adding and subtracting counter, a small 110-v. lamp illuminating the counter, a yellow-domed 6-v. lamp beneath the counter, and switches for resetting the counter. The yellow lamp was turned on at the beginning of each trial to signal *S* that a new trial had begun; when it went off, he was to make a choice of one of the four switches before him.

The equipment outside of the booth consisted largely of a timing motor, Western Union tape transmitters, a number of relay memory circuits, and lamps on a panel which permitted *E* to record *S*'s responses and the results of these responses (in terms of wins and losses). There were also a number of rotary switches which permitted *E* to change the correspondence between channels of the Western Union tape transmitters and *S*'s response alternatives without typing new tapes.

Conditions.—There were four experimental conditions. These can best be described by matrices of conditional probabilities. These matrices give the probabilities of a given response being correct on Trial n when it is assumed that a particular response occurred on Trial n-1. In the following matrix let p_{ij} be the probability that making Response i will lead to a win (add a point to the counter) if Response j had been made on the preceding trial, Trial n-1. Specifically in the matrix below, p_{31} would be the probability that Response 3 will be reinforced if Response 1 had just been made on the preceding trial.

Responses Available on Trial n	Response Chosen on Trial $n-1$			
	1	2	3	4
1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	p_{22}	p_{23}	p_{24}
3	p_{31}	p_{32}	p_{33}	p_{34}
4	p_{41}	p_{42}	p_{43}	p_{44}

In the preceding matrix, response alternatives designated by the numbers 1 to 4 may be interpreted as a particular switch for a particular *S*. Since, however, this correspondence between the responses numbered 1 to 4 and the switches were permuted for different *S*s, no physical interpretation of the numbered responses as the one on the extreme right, etc., is possible. Instead, the number designations only represent some *fixed* alternative for a given *S*.

TABLE 1

MATRICES GIVING THE PROBABILITY OF A RESPONSE LEADING TO A WIN AS DEPENDENT UPON THE PRECEDING RESPONSE

Responses Available on Trial n	Response Chosen on Trial $n-1$											
	1	2	3	4		1	2	3	4			
	Condition A					Condition B						
1	.25	.25	.25	.25		.00	.33	.33	.33			
2	.25	.25	.25	.25		.33	.00	.33	.33			
3	.25	.25	.25	.25		.33	.33	.00	.33			
4	.25	.25	.25	.25		.33	.33	.33	.00			
	Condition C					Condition D						
1	.00	.50	.50	.00		.00	1.0	.00	.00			
2	.00	.00	.50	.50		.00	.00	1.0	.00			
3	.50	.00	.00	.50		.00	.00	.00	1.0			
4	.50	.50	.00	.00		1.0	.00	.00	.00			

The four conditions in this experiment are given by the matrices in Table 1. Here the p_{ij}'s are replaced by specific probabilities.

Condition A was a control condition in which nothing *S* did could alter his average probability of winning. Thus, regardless of how frequently *S* repeated the same response or changed his response on successive trials, he maintained the same expected probability of success. In Cond. B, *S* could only increase his probability of success by never repeating the same response. In Cond. C, for any given choice on Trial n-1, only two of the remaining three alternatives had a positive probability of leading to a win. The probability of each of these responses leading to a win was .5. In order to maximize their probability of winning, *S*s had to (*a*) learn to avoid repeating the same response, and (*b*) learn to avoid one other of the remaining alternatives, where the incorrect alternative depended upon the preceding response. In Cond. D, for each alternative on Trial n-1 there was one, and only one, alternative which had a probability greater than zero of leading to a win on Trial n. The probability of this alternative was 1.0.

Procedure.—The 40 *S*s in this experiment were all from elementary psychology classes at the University of Wisconsin. Ten *S*s served in each condition. The same instructions were given to all *S*s in all conditions. When *S* reported to the laboratory, he was immediately seated in the booth. All settings of the apparatus determin-

ing the condition under which *S* was to serve had been made prior to his arrival. When *S* was seated, instructions which described the "game" were given. Both *S* and the machine, *S* was told, had a choice of one of four alternatives on each trial. Then, depending upon the combination of choices made by both *S* and the machine, either *S* or the machine won. If *S* won, a counter added a point; if *S* lost, this counter would subtract a point. All *S*s were given 1000 points at the beginning of the experiment. Each *S* was further instructed that the machine in many ways was built to act like a person. Consequently, *S* was informed, the machine would do different things depending upon what *S* had previously done. That is, *S*s were told that the machine had a memory and would use this memory to change its behavior. In addition to these instructions, *S*s were told to choose only one switch on each trial, and to pull this switch all the way back to insure good contact.

After a brief demonstration each *S* was given 402 trials; each response was considered as one trial. The first trial was not counted, and the last trial was added to round out to an even 400 the number of first-order conditional probabilities. Each trial was of 4-sec. duration. At the end of each trial the counter either added or subtracted a point.

Results

The results are presented principally as mean probabilities in Fig. 1 and 2. These means are computed for successive 100-trial blocks. These means for each block of trials are partially ordered by the Revised Duncan Range Test (2) and the Paired Replicates Test (7). The first was used when there were three or more independent means to be compared, the last when there were two dependent means to be compared. When results are spoken of as significant, the .05 level will always be employed. For Fig. 1 and 2, the points contained within an enclosed line differ significantly from all those points outside of the enclosure, but do not differ significantly among themselves.

Probability of response change.— One of the results of primary interest in this study concerns the effects of the different conditions upon the probability of a change in response on successive trials. The partial orderings of such probabilities, by blocks of 100 trials, are given in Fig. 1. It is seen that *S*s under all conditions tended to avoid repetition. For both Cond. C and D the number of repetitions differ significantly from chance for all four blocks of trials. Repetitions under Cond. B differ significantly from a chance level on Blocks 2, 3, and 4. For Cond. A (the control condition), repetitions are significantly fewer than would be expected by chance on the second and last block of trials.

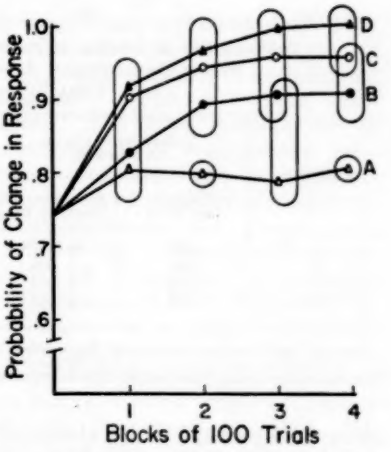

Fig. 1. Probability of a change of response on successive trials.

The probability of repetition leading to a win for both Cond. D and B is zero. By the fourth block of trials, however, the probability of a change for Cond. D is significantly greater than that for Cond. B. The probability of change for each of these conditions is given in Table 2.

Following each response, one of two events occurred: either *S* won, or he lost. It is possible that these events

TABLE 2

PROBABILITIES OF CHANGE OF RESPONSE AND PROBABILITIES OF RESPONDING TO THE CONDITIONALLY REINFORCED ALTERNATIVES FOR THE FINAL BLOCK OF 100 TRIALS

Condition	Probability of Change of Response		Probability of Responding to Conditionally Reinforced Alternatives		
	Value for Last Block	Difference from Preceding Block	Value for Last Block	Difference from Preceding Block	Average Value for Last Block
A	.801	+.017[a]	1.000[b]	+.000[b]	.250[b]
B	.906	+.005	.906	+.005	.302
C	.953	+.001	.678	−.010	.339
D	.999	+.004	.995	+.009	.995

[a] A "+" represents an increase over the preceding block of trials.
[b] As *S* must respond to one of the four alternatives, and each one has an equal probability of leading to a win, these values are a priori values rather than empirical ones.

affected the probability of a change of response on the next trial. In Fig. 2 are presented graphs of the probability of a change in response after a win and after a loss for each condition separately. The curves labeled W are the curves representing the probability of a change in response after a win, and the curves labeled L are those giving the probability of a change in response after a loss. It is seen that for Cond. A and C there are no differences for any of the blocks of trials. Initially there are no differences for Cond. B. Then for the last 200 trials the curves diverge; there is a significantly greater probability of a change in response after winning as compared with a change after losing. For Cond. D, the curves are initially different in the same direction as in Cond. A, but later converge. Furthermore, when the probability of a change after a loss for Cond. B is compared with the probability of a change after a loss for Cond. A, they do not differ significantly for any of the four blocks of trials. This result was not true for changes in responses after wins.

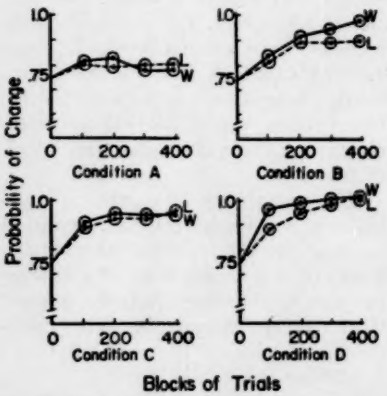

FIG. 2. Probability of a change of response on successive trials following a win and following a loss.

Probability of response to conditionally reinforced alternatives.—By a conditionally reinforced alternative is meant a switch which has a fixed positive probability of leading to a win. This probability is contingent upon the immediately preceding response. Thus, a conditionally reinforced alternative is defined with reference to a fixed probability. For example, in Cond. C, after Response 3, Response 1 is a conditionally reinforced alternative with a probability of .5. The probability associated with any given switch will, however, vary with differences in the preceding

response. Hence, a conditionally reinforced alternative cannot be identified with a physical location.

The results concerning the probability of a change have already indicated that *S*s did not always respond proportionately to the probability of winning. Table 2 gives the average probability of responding to a conditionally reinforced alternative during the last block of trials. The location of the conditionally reinforced alternatives, even for the same preceding response, was varied from *S* to *S* to control any position bias which might exist. It is thus not meaningful to give probabilities for each of the alternatives separately. Only an average for all alternatives with the same probability can be given. Examination of individual records revealed no strong preferences between conditionally reinforced alternatives. That *S*s were very close to asymptote by the fourth block of trials (under the assumption of a monotone singly inflected acquisition curve) is indicated by the small difference between Blocks 3 and 4 in Table 2.

It is seen from Table 2 that the probabilities of reinforcement are rather closely matched by the average relative frequencies of response to an alternative in every condition except Cond. C. Here it is seen that the probability of responding to a conditionally reinforced alternative is considerably lower than the probability with which that alternative will lead to wins. This discrepancy seems to be largely due to *S*'s inability to discriminate between the three nonrepeating alternatives (the three alternatives which do not involve repeating a response). The probability of responding to the alternative (other than the repeating alternative) which never led to a win was .321, .302, .263, and .277 for Blocks 1 through 4, respectively. Thus, at the end of 400 trials *S*s were still responding with a greater than chance frequency to a conditional alternative which was never reinforced. The final probabilities of responding to a conditionally reinforced alternative are, however, ordered by the probability of conditional reinforcement.

Probability of winning.—The probability of *S*'s winning can be essentially computed from his probability of responding to the conditionally reinforced alternatives. The highest probability of winning selected from all blocks of trials were for Cond. A through D, respectively, .250, .301, .333, and .999. Thus, only in Cond. D was *S* able to win a good deal more than he could win by playing randomly.

Discussion

The two results of most interest in this study concern (*a*) the effects of the conditional probabilities of reinforcement upon the probability of a change in response, and (*b*) the differences between response probabilities and the probabilities of conditional reinforcement.

Probability of a change in response.—The results clearly show that not reinforcing the repetition of a response leads to a decrease in repetition. The probability of a change in response on successive trials for Cond. B, C, and D was found to be generally greater than that for Cond. A. This decrease in repetition, however, was not only dependent upon the zero probability of reinforcement for repetition. The *S*s in Cond. D showed the most rapid decrease in repetition. In this condition, where a probability of 1.0 was associated with a single alternative, *S*s by responding to this alternative could win much more often by changing their response than by any pattern of responding in Cond. B and C. Thus, although repetition was not reinforced in two of the other three conditions, no pattern of changing responses would lead to as high a prob-

ability of winning in these conditions. Thus, the distribution of probabilities on other than the repeating alternatives, as well as the sum of probabilities on these alternatives, must be taken into consideration.

It was further found that the probability of a change in response was never greater after a loss than after a win, and sometimes greater after a win. In order to have been reinforced, however, *S* must have changed his response. Thus, the probability of a change of response following a change of response was greater than following a repetition. This, together with the quickly attained high probability of a change for Cond. D, indicates that a high probability of a change of response must be rather constantly reinforced in order to be maintained. Further evidence in this direction is given by the lack of convergence of a change after wins and a change after loss curves for Cond. B in Fig. 2. It thus appears that learning not to repeat is highly dependent upon making specific other responses which are reinforced. That is, of course, congruent with the competition of response theories of extinction.

It should be noted that the precise effect of dividing the probabilities of reinforcement between two or three responses has not been stated. This is because *S*'s sequential behavior has not been analyzed beyond first-order conditional probabilities. This current experiment could be viewed as a generalization of rote serial learning conditions. In Cond. D it was possible for *S* to learn a simple invariant repeating series. In Cond. C and B, *S* could learn any of several permissible response chains, some of which would have more than the four links of the response chain in Cond. D. A more extensive analysis of response sequences would be required to differentiate between the effects of increasing the number of permissible sequences, and the effects of decreasing the probability of a *change* in response leading to a win.

A decreased probability of changing alternatives on successive trials thus could be one of the reasons for the disagreement between probability learning experiments involving two, and more than two, responses. Assume that the maximum probability associated with an alternative remains fixed. Then, as the number of permissible alternative responses increases, the probability of a change in response being reinforced decreases. As this experiment has shown, it is not only the sum of the probabilities on opposing alternatives that is important, but also the distribution of these probabilities on the opposing alternatives. It is the latter which affects more directly the probability of a change in response leading to a reinforcement.

Response to conditional relative frequencies.—In Cond. D, almost every *S* was responding perfectly by the end of the experiment. With conditional relative frequencies of less than one, however, *S*s matched the conditional relative frequencies less well. The preceding discussion has already indicated that even though *S*s were probably quite close to asymptote, they were still choosing alternatives with zero probabilities of reinforcement. Further, for the fourth block of 100 trials in Cond. C, *S*s were responding to an alternative with zero probability (which was not the repeating alternative) with a relative frequency of .277. Such a failure to discriminate conditional alternatives having a zero probability of reinforcement means that *S*s could not match those alternatives with a probability greater than zero. Thus, on alternatives having a greater-than-zero probability of reinforcement, *S*s consistently responded with a probability smaller than the probability of reinforcement.

The *S*s did, however, in Cond. B and C tend to divide their responses equally between the conditionally reinforced alternatives. In Cond. B, then, the relative frequencies of responding to conditionally reinforced alternatives matched fairly closely the probability of reinforcement. In Cond. C, *S*s matched the probability of reinforcement less well. This was largely due to the previously

discussed failure to differentiate between the nonrepeating alternatives.

This experiment does not represent a very comprehensive test of the general hypothesis that *S*s will tend to match relative frequencies of reinforcement with relative frequencies of response. It does, however, show that in certain limiting cases, an extension of this hypothesis is not verified.

The situation as a social situation.—Because of the dependency of events following a response upon the prior response, this experimental situation may be regarded as an elementary social situation. In this regard, this study contains two implications for the study of social behavior. First, even very simple contingencies, such as those in Cond. C, are learned slowly and inadequately. If this is an unbiased estimate of learning ability in partially reinforced contingent situations, it is clear that most responses in social situations must be made on the basis of concepts and stereotypes. This reliance upon concepts for changes in behavior which lead to quick adjustments to uncertainties in everyday situations has been developed in a recent paper by Fitts (**4**).

Secondly, when in a social situation a response has a very low probability of being repeated, a possible reason for this could be that a change in response has a high probability of being reinforced. Certain forms of "unstable" behavior could possibly be produced in this way.

Summary

An experiment was conducted in which conditional probabilities of responses leading to reinforcement were associated with preceding responses. These contingent probabilities were such as to never, except in a control condition, lead to a reinforcement for the repetition of a response. The results showed that *S*s in all cases, including the control condition, tended to change alternatives on successive trials more frequently than would occur by chance. Probabilities of changing responses on successive trials increased most rapidly when the probability of being reinforced on a single other alternative was greatest. Largely through continued response to alternatives which had an associated zero probability of leading to reinforcement, certain exceptions to the probability-matching hypothesis were found to hold.

References

1. Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–104.
2. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.
3. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of a statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
4. Fitts, P. M. Toward a general theory of human information processing. Poughkeepsie, N. Y.: IBM Res. Rep., 1957 (in press).
5. Gardner, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.
6. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
7. Wilcoxon, F. Some rapid approximate statistical procedures. New York: Amer. Cyanamid Co., 1949.

(Received October 21, 1957)

Journal of Experimental Psychology
Vol. 56, No. 5, 1958

STIMULUS SPACING AND THE JUDGMENT OF LOUDNESS [1]

JOSEPH C. STEVENS

Harvard University

Under some circumstances the spacing of auditory stimuli on the intensity scale may have a strong effect on judgments of loudness. Under other circumstances, however, the judgment of loudness seems to be independent of the "context" provided by stimulus spacing.

Effects attributable to the spacing of the stimulus ensemble may be especially marked when the method of constant stimuli is used to determine the value of a stimulus judged to be half as loud as a given standard (**1, 6**). The value judged half depends in a sensitive manner on how the "variable" stimuli are distributed relative to the standard.

A similar dependence on spacing is observed when listeners try to judge apparent loudness on a category rating scale (i.e., on a scale of equal-appearing intervals). In this case the curvature of the category scale is systematically related to the relative spacing of the stimuli (**5**).

On the other hand, there is evidence that under the method of magnitude estimation the judgment of loudness may be relatively independent of stimulus spacing (**4**). This procedure calls for a direct estimation of the apparent loudness of each of a series of stimuli. The *O* tries to assign numbers proportional to the apparent loudnesses, and he is free to use any numbers he thinks appropriate.

[1] This research was carried out under Contract Nonr-1866(15) between Harvard University and the Office of Naval Research, U. S. Navy (Project Nr142-201, Report PNR-208). Reproduction for any purpose of the U. S. Government is permitted.

Since it is an important question whether the form of the loudness scale determined by magnitude estimation is independent of stimulus spacing, two experiments were undertaken. The first experiment compared the magnitude estimations of stimuli spaced uniformly in decibels with those of stimuli spaced uniformly in loudness (sones). The same spacings were also used to generate category scales to which the scale based on magnitude estimation could be compared. The second experiment investigated the extent to which local perturbations in the loudness function can be brought about by an extreme bunching of the stimuli in narrow regions of the total stimulus range.

Experiment I

Method.—The stimuli for Exp. I and II were produced by a white noise generator and delivered through PDR-8 earphones mounted in MX-41/AR cushions. A low-pass filter with a cut-off at 6000 cps limited the bandwidth of the noise. In Exp. I the *O* pressed in sequence two keys, one to present a standard loudness before each variable, the other to present the variables he was to judge.

The *E* assigned a number to the loudness of the standard noise and instructed *O* to estimate the loudness of variable noises relative to the standard noise. He was told to make his numbers proportional to apparent loudness, e.g., if the standard was "100" and the variable sounded one-fourth as loud, he was to call it "25."

In the first part of the experiment 10 *O*s estimated the loudness of nine variable stimuli, which were spaced at intervals of 5 db and presented in a different, irregular order to each *O*. Half the *O*s began by estimating each variable twice against a standard at 90 db re .0002 dyne/cm.^{2}, called "100." After a short rest they received a new standard, 45 db, called "1," against which they again estimated each variable twice.

The other *Os* began with the 45-db standard, called "1," but otherwise followed the same procedure.

A couple of months later the same 10 *Os* were recruited for another session. Their task was the same, but the stimuli were spaced in a different way: the values were set to correspond roughly to uniform intervals of apparent loudness, as inferred from the results of the first part of the experiment. In all other respects the two parts of the experiment were the same.

In still another part of the experiment, 10 *Os* judged the loudness of the white noises on a 7-point category scale. Five *Os* first judged the noises spaced at equal decibel intervals and then a day or two later judged the noises spaced at approximately equal loudness intervals. The other five *Os* followed the reverse procedure. The instructions were to assign each noise to one of seven equally spaced categories designated by the numbers "1" through "7." The loudest and softest noises in the series were identified as such before the session began.

Results.—The geometric means of the loudness estimates of each stimulus appear in the log-log coordinates of Fig. 1. The data for the two types of stimulus spacing are plotted together for easy comparisons. The two straight lines have the slope of the loudness function for a 1000-cycle tone (.3) and appear to fit the estimates for noise quite well. The reason for using the geometric mean as a measure of location is that an examination of data reported earlier (**2**) shows that the medians and geometric means of the loudness estimates of 70 *Os* did not differ significantly, and that to a first approximation the *Os*' estimates of a given stimulus were normalized by logarithmic transformation.

If the spacing of the stimuli has a systematic effect on estimates of loudness, it is not obvious from this investigation. At least it does not appear that spacing in uniform loudness intervals and spacing in equal decibel steps produce any notable differences.

Quite different, however, is the effect of these spacings on the category rating scales obtained with the two sets of stimuli. Figure 2 shows the mean category assignments plotted against apparent loudness. (The abscissa scale is proportional to loudness in sones.) The results agree with earlier studies in that the category judgments are not linear, but are concave downward when plotted against apparent loudness.[2] Moreover, the form of the function depends, as anticipated, on the spacing of the stimuli. Spacing by nearly equal loudness intervals produced less departure from

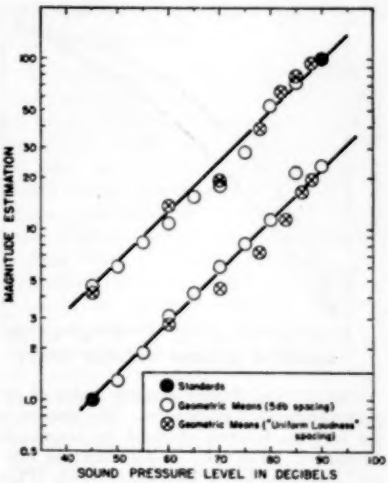

FIG. 1. Loudness estimation under two kinds of stimulus spacing. The open circles stand for the geometric means of estimates of stimuli that were spaced at 5-db intervals. The crossed circles stand for the geometric means of the estimates of stimuli spaced at roughly equal loudness intervals. Each point represents 20 estimates (10 *Os*). A standard at 45 db, called "1," was used to obtain the data for the lower curve; a standard at 90 db, called "100," was used to obtain the data for the upper curve. The straight lines drawn through the standards have the slope (.3) of the sone scale.

[2] For a discussion of the reasons for this nonlinearity see Stevens and Galanter (**5**).

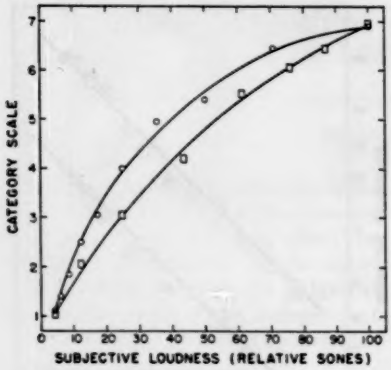

FIG. 2. Two 7-point category scales of apparent loudness of white noise. The abscissa is the loudness function obtained by magnitude estimation. (The value of 100 sones was arbitrarily assigned to the loudness of 90 db SPL.) The circles stand for the mean category number assigned to stimuli spaced at equal decibel intervals. The squares stand for the mean category number assigned to stimuli spaced at nearly equal loudness intervals.

linearity than spacing by equal decibel intervals.

The over-all results of Exp. I demonstrate that, as regards the effect of stimulus spacing, magnitude estimation is a more stable process than category rating. The question is, How stable is it? How extreme would the bunching and spreading of the stimuli need to be in order to produce an effect? In Exp. II a deliberate attempt was made to distort the loudness function by an extreme kind of spacing.

EXPERIMENT II

Method.—The stimuli, again white noises, were spaced in four ways: (*a*) at equal decibel intervals, (*b*) bunched together toward the high end of the range, (*c*) bunched toward the low end, and (*d*) bunched at the middle. That the bunchings were extreme is apparent from the following lists of sound pressure levels of the stimuli:

Spacing *a*: 40, 50, 60, 70, 80, 90, 100 db

Spacing *b*: 40, 66, 80, 90, 92, 94, 98, 100 db

Spacing *c*: 40, 42, 44, 46, 50, 60, 70, 80, 100 db

Spacing *d*: 40, 50, 62, 64, 66, 68, 70, 80, 90, 100 db

All four parts of the experiment employed the same range (40 to 100 db re .0002 dyne/cm.^{2}), the same standard (80 db, called "10"), and the same instructions. Fourteen *O*s participated in each part.

The *O* was instructed to press a key that delivered the stimuli and to assign numbers proportional to the apparent loudness. He was told that the loudness of the first stimulus presented (80 db) should be called "10."

Although this standard came on as a standard only at the beginning of a series, *E* presented the same intensity again along with the other variables. After the first series, the standard (identified as such) came on again, followed by the same set of variables in a different order. The order of the stimuli was also varied from one *O* to another, and was always irregular.

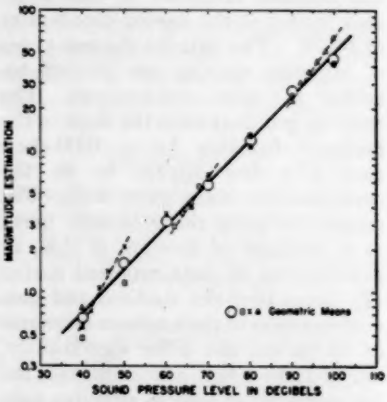

FIG. 3. Loudness estimation under four kinds of stimulus spacing. The circles stand for the geometric means of the estimates of stimuli spaced at 10-db intervals. The straight line, whose slope is .308, was fitted to these points by the method of least squares. The other points represent the geometric means of the estimates of stimuli that were bunched at various regions of the total range. Each point represents the geometric means of 28 estimates (14 *O*s). The same standard (80 db, called "10") was used for each condition of stimulus spacing, and it was presented only at the beginning of a series of estimates.

Results.—Figure 3 records the results. The solid straight line was fitted by the method of least squares to the geometric means of the estimates of only those stimuli that were spaced at equal decibel intervals. The slope of the line came out close to .3; that is to say, 2:1 ratios of loudness correspond approximately to 10:1 ratios of intensity. As noted before, this slope coincides with that of the loudness function for a 1000-cycle tone. The function confirms the result of earlier experiments (cf. **2, 3**) which indicated that above about 40 db the apparent loudness of white noise approximates a power function of stimulus intensity, and that the slope of the function roughly equals that for a 1000-cycle tone.

The geometric means of all the other estimates also appear in Fig. 3. Despite the bunching of the stimuli into narrow regions of the total range, the points deviate but little from the solid line. Admittedly the extreme spacing had an influence on the estimates of loudness, but the influence is surprisingly minor. As shown by the dotted lines the deviations that occurred are largely confined to the region of bunching.

It may be seen in Fig. 2 that bunching may increase the *local* slope of the function, while leaving the over-all slope relatively unaffected. The *O*s tended, in other words, to exaggerate the differences among closely spaced stimuli. On those occasions when a stimulus differed only slightly—say, 2 db—from the preceding one, *O*'s estimate might well represent two intentions on his part: he tries (*a*) to estimate the apparent loudness of the stimulus, but also (*b*) to communicate the fact that he heard a difference between stimuli of nearly equal loudness. In other words, a tendency to concentrate on loudness differences might explain the deviations brought about by bunching. Whatever the true explanation for the local changes may be—and perhaps any explanation ex post facto might seem gratuitous—the important thing to note is how little the function as a whole is changed by stimulus bunching.

The bunching of the stimuli also had an effect on the variability of the estimates. *SD*'s of the logarithmically transformed estimates appear in Fig. 4 as a function of proximity of the variable stimulus to the standard. Although the *SD*'s look roughly constant at about .2

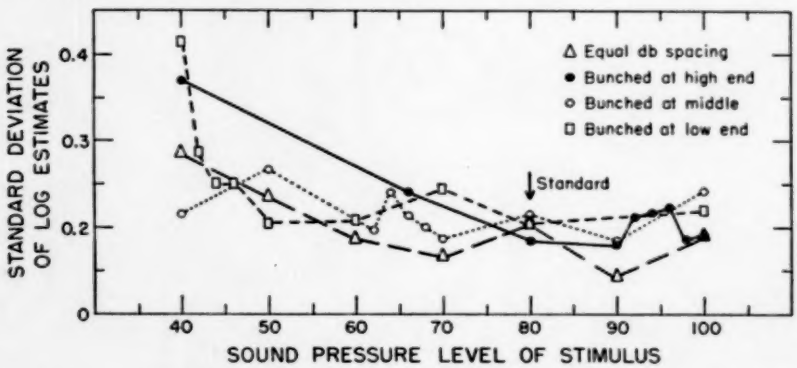

FIG. 4 The variability of estimates as a function of proximity of the variable stimulus to the standard. The ordinate is the standard deviation of the logarithmically transformed estimates obtained in Exp. II. The abscissa is the SPL of the stimulus.

log unit, the variability is somewhat larger at points far removed from the standard and at regions of stimulus bunching.

In contrast to category scales, whose form is sensitive to the spacing of the stimuli, ratio scales established by magnitude estimation seem to be quite stable. It would appear that even very unusual spacing produces only second-order changes in the magnitude estimates. This finding may be regarded as a significant recommendation for the use of ratio scaling procedures, as opposed to category scaling procedures, especially in view of the fact that category judgments are generally inconsistent with ratio scales of apparent magnitude. The relative insensitivity of magnitude estimates to stimulus spacing also stands in marked contrast to the finding that the method of constant stimulus yields judgments of one-half loudness that are radically contingent upon the context of stimuli available to the observer.

Summary

The scaling of loudness by the method of magnitude estimation under a number of different stimulus spacings confirms the finding that the loudness of white noise is essentially a power function of stimulus intensity above about 40 db SPL. The over-all form and slope of the function turn out to be relatively insensitive to the spacing of the stimuli. Small local distortions can be brought about, however, by bunching the stimuli into narrow regions of the total range. In contrast to the stability of magnitude estimates, category judgments of loudness vary substantially with the spacing of the stimuli. The results argue in favor of magnitude estimation over category rating in the scaling of apparent magnitude.

References

1. Garner, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.
2. Stevens, J. C., & Tulving, E. Estimations of loudness by a group of untrained observers. *Amer. J. Psychol.*, 1957, **70**, 600–605.
3. Stevens, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
4. Stevens, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.
5. Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.
6. Stevens, S. S., Rogers, M. S., & Herrnstein, R. J. The apparent reduction of loudness: A repeat experiment. *J. acoust. Soc. Amer.*, 1955, **27**, 326–328.

(Received October 28, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

RESISTANCE TO EXTINCTION AS A FUNCTION OF CONTINUOUS OR INTERMITTENT PRESENTATION OF A TRAINING CUE [1]

MELVIN H. MARX

University of Missouri

The present experiment was designed to test the prediction that resistance to extinction would be greater for intermittent than for continuous presentation of a cue that had been continuously paired with reinforcement during training. This "partial extinction" procedure was devised to test the more general theoretical assumption that maintenance of motivation during extinction would be greater with the irregular than with the normal consistent schedule. The rationale underlying this assumption is that extinction itself is primarily a function of the loss of the specific motivating properties of the stimulus complex.

METHOD

Subjects

The *S*s were 27 naive male albino rats, approximately 100 days old at the start of the experiment. There were 14 experimental and 13 control *S*s. The data obtained from one additional control *S* were eliminated because of an error in feeding schedule maintenance on the day prior to extinction.

Apparatus

A simple straight runway was used. It had a 1-ft. starting box, a 7-ft. stem, and a 1½-ft. goal box. Alley width was 3 in. in the starting box and stem, 10½ in. in the goal box. Alley height was 5 in. throughout. The runway was covered with hardware cloth. The inside of the alley was Masonite in the starting box and stem; the goal box was also Masonite but could be covered with black insert boards.

Starting time, running time, and time in the goal box were all measured automatically by means of 3-in. electronic contact switches set in the floor and connected to Standard Electric timers. Guillotine doors, operated manually, were placed at the end of the starting box (6½ in. in front of the first floor switch), and at the beginning of the goal box (3 in. after the second floor switch). An intermediate door was placed 10½ in. in front of the goal-box door, to prevent retracing in the stem.

A glass sponge cup (3½ in. in diameter, 1¾ in. high) was used as a food cup. It was placed at the end of the goal box, behind a black shield which prevented *S* from seeing whether or not the food cup was in place without stepping on the third floor switch (placed 4 in. from the end of the goal box). This switch gave a measure of running time in the goal box.

Experimental Design

All *S*s were given the same training, running to the food cup on every trial (100% reinforcement schedule). Experimental *S*s were then extinguished with the empty food cup present on only 50% of the trials; control *S*s were extinguished in the usual way, with the empty food cup present on every trial.

Procedure

Deprivation and accommodation.—All *S*s were placed on a 24-hr. feeding program, being allowed 1 hr. of feeding on Little Friskies pellets each day, for five days prior to accommodation. Two days of accommodation to the runway were then given. On Day 1, the two *S*s that were caged together were placed first in a neutral (tempered Masonite) goal box and allowed to explore the empty goal box and runway for 10 min. After 5 min. in individual detention cages, the two *S*s were placed in the closed-off goal box and allowed to eat 10 Little Friskies pellets (total weight about .5 gm.). Each *S* was then placed in the runway individually at points progressively

[1] This investigation was supported in part by a research grant, M-817, from the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.

farther from the goal box and allowed to eat two pellets. Two trials were given at each point. All trials were separated by 20-sec. periods in the detention cage. When *S* ran from the starting box to the goal box within 90 sec. on two consecutive trials, his accommodation was considered complete. If on any trial this time limit was exceeded, *S* was placed in the goal box directly and allowed to eat. The next trial was given from the point of the last successful trial. On Day 2 this procedure was repeated, except that only one trial was given from each of the six starting points.

Training.—The next six days were training days, during each of which all *S*s were given six trials to two food pellets in the cup. The starting door was raised from 3 to 6 sec. after placement of *S* in the box. The *S* was allowed 10 sec. to traverse the stem and 10 sec. in the black goal box to eat. If *S* failed to run to the goal box within 10 sec., he was placed at the food cup by *E* and allowed 10 sec. to eat. An intertrial interval of 20 sec. was used, with *S* kept in the detention cage. Following the sixth trial, *S* was kept for 5 min. in the detention cage and then given 1 hr. of feeding on Little Friskies pellets. This feeding procedure, which had also been used during the two accommodation days, resulted in an effective drive of approximately 22 hr. at the start of each training session.

Extinction.—The *S*s were randomly divided into the experimental and control groups. Experimental *S*s ran to the empty food cup behind the black shield on 50% of the trials. The schedule was determined by a table of random numbers with the restriction that no run of more than three successive cup or no-cup trials was permitted. A different schedule was used for each *S*. Control *S*s ran always to the empty food cup.

Extinction trials were massed. All *S*s ran a minimum of 36 trials, with the same 20-sec. interval in the detention cage as had been used in training. The extinction criterion was failure to complete a run within 180 sec. from opening of the starting door on two consecutive trials.

Results

Trials to extinction.—The mean trials to extinction of the alley response (to two successive 180-sec. trials) were 110.43 and 83.00 for the experimental and control groups, respectively. Significantly greater resistance to runway extinction was shown ($F = 9.42$, $P < .01$), as was predicted, by the experimental *S*s. The mean trials to extinction of the goal-box response (to two successive 10-sec. trials) were 45.29 and 34.00 for the experimental and control groups, respectively. This difference, although in the same direction as that for the alley response, was short of significance at the .05 level ($F = 3.01$, $P < .10$). The criterion for the extinction of the goal-box response was met by all *S*s before the criterion of runway extinction.

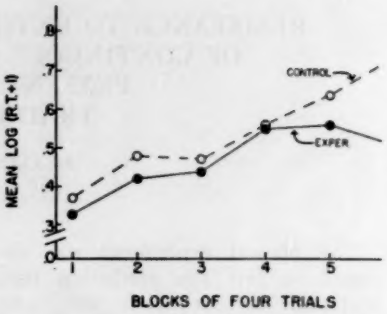

FIG. 1. Mean log running time + 1.00 for the first 24 trials of extinction.

Running times.—Mean running times for the first 24 trials of extinction are shown in Fig. 1. Running times were transformed to log (R.T. + 1.00). They are plotted in terms of 6 blocks of 4 trials, in order to show more clearly the trends present in the analysis, which was done on a trial-by-trial basis. This number of trials was used in order to avoid inclusion of any 180-sec. extinction scores (the first such occurring on Trial 25 for one of the control *S*s).

The more rapid extinction of the control *S*s is again evident by this measure. The analysis of variance is shown in Table 1. Although the over-all group difference missed significance at the .05 level, the performance levels clearly diverged over

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE FOR LOG RUNNING TIME + 1.00 FOR THE FIRST 24 TRIALS OF EXTINCTION

Source	*df*	*MS*	*F*	*P*
Between *Ss*	26			
Groups (G)	1	.8886	3.52	<.08
Error	25	.2522		
Within *Ss*	621			
Trials (T)	23	.2900	5.50	<.001
T × G	23	.0928	1.76	<.025
Error	575	.0527		
Total	647			

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR 1/GOAL-TIME FOR THE FIRST 24 TRIALS OF EXTINCTION

Source	*df*	*MS*	*F*	*P*
Between *Ss*	26			
Groups (G)	1	4.5781	5.05	<.05
Error	25	.9059		
Within *Ss*	621			
Trials (T)	23	.8609	5.28	.001
T × G	23	.1009	0.62	
Error	575	.1631		
Total	647			

trials, as shown in Fig. 1 and indicated by the significant Group × Trial interaction.

The groups had reached approximately the same asymptote by the fifth training day ($F < 1.00$ between groups for the last 12 training trials).

The running-time scores for the first 12 cup trials in the experimental group were compared with the first 12 cup trials in the control group. No difference was found (each group superior on half of the trials, $F < 1.00$), in spite of the fact that 12 additional no-cup trials were interpolated for the experimental *Ss*.

Goal times.—The approach times within the goal box (running time for the 18-in. segment) are shown in Fig. 2. The consistently faster running of the experimental *Ss* was clearly significant, as is shown in Table 2. The groups had reached approximately the same asymptote by the fifth training day ($F = 1.93$ between groups, 4.24 required at .05 level).

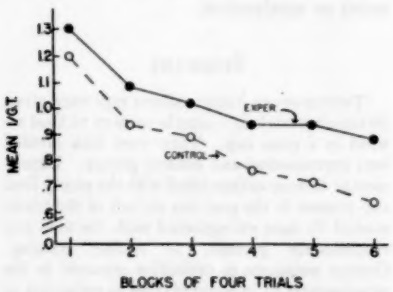

FIG. 2. Mean reciprocal of goal time for the first 24 trials of extinction.

Starting times.—Although the starting times for the experimental *Ss* were shorter than those for the control *Ss*, an analysis of variance of these measures (1/S. T.) indicated that the difference fell short of significance when computed for the first 36 trials of extinction ($F = 2.80$, $P > .05$, for 1, 25 df.).

Cup vs. no-cup trials.—Since the trials-to-extinction and the running-time measures showed a reliable difference in favor of the experimental group, the next question concerned the locus of the superiority. In other words, was the faster and more persistent running found in extinction for the cup-no-cup *Ss* a function of faster running primarily following the exposure to a cup? Or was the effect a general one, with both cup and no-cup trials producing essentially the same running times? To answer this question some direct comparisons of cup and no-cup scores were made in the experimental group.

Running times were computed for

the first 12 trials immediately *following* a cup and compared with the comparable scores immediately following a no-cup condition. The scores of each *S* were compared in order to classify him as a cup-faster or a no-cup-faster *S*. Ten of the 14 experimental *S*s ran faster following the cup condition. If the one-tailed test is used, because the direction of this difference was predicted in advance, the difference may be regarded as significant at the .057 point ($\chi^2 = 2.571$) by a linear interpolation.

Discussion

The obtained results clearly support the prediction of greater resistance to extinction for the "partial extinction" condition. This is true of both running times but not of the less sensitive starting-time measure. The intermittent presentation procedure resulted in almost exactly the same alley-running extinction for the first twelve cup presentations in both groups.

These results cannot be easily explained on the basis of a simple application of some major theories of extinction. The so-called "discrimination hypothesis" (1, 2, 5) is clearly contradicted. The control group, with the greatest similarity between training and test conditions (always a cup present), showed the least resistance to extinction. The work-inhibition interpretation (4) has no clear basis for predicting the difference between the groups, since there is no evidence of a difference in work or effortfulness during the test. Finally, the group that would presumably receive the most secondary reinforcement from the cup within any given block of test trials (such as the 24 used to measure running time) would be the control group, since it had twice as many occurrences of the secondary-reinforcing cue.

An alternative interpretation involving secondary reinforcement would consider the rate of the loss of power, or extinction, of a secondary-reinforcing cue. From this point of view, the intermittent-presentation schedule simply delays the extinction of such a cue, and the total number of presentations is the important factor. This position is similar to inferences that may be drawn from Estes' (3) statistical-association theory. Even from this point of view, however, it is of interest that the intermittent-presentation schedule permits the interpolation, as mentioned above, of an equal number of no-cup trials without appreciable loss of strength of the running response.

The interpretation that is most strongly favored by the results, in my opinion, is one that holds extinction to be, fundamentally, a function of the degree to which motivation to make the specific response is maintained. Loss of such motivation would then occur at a more rapid rate under the continuous as compared with the intermittent schedule. The difference between this interpretation and the one mentioned directly above may be, at heart, one of semantics —partly revolving around whether the concept of secondary reinforcement or that of secondary motivation is used. Further, more specifically designed research will be needed to clarify this issue.

The suggestion that more experimental *S*s ran faster immediately following a cup trial as compared with a no-cup trial in extinction is of some interest. If this tendency can be replicated it would seem to support the general interpretation of the cup, in the cup-no-cup context, as the source of the secondary effect—reinforcement or motivation.

Summary

Twenty-seven hungry albino rats were given 36 training trials in a simple runway to food reward in a glass cup. They were then divided into experimental and control groups. Experimental *S*s were extinguished with the empty food cup present in the goal box on half of the trials; control *S*s were extinguished with the food cup continuously present, as during training. Greater resistance to extinction occurred in the experimental *S*s, in terms of trials to extinction as well as running times. They also ran significantly faster within the goal-box segment.

The results are discussed from several points of view. It is concluded that they support the original assumption that resistance to extinction is a function of the maintenance of motivation, which is facilitated by the intermittent schedule on which the food-cup cue is introduced during extinction.

References

1. Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.
2. Elam, C. B., Tyler, D. W., & Bitterman, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. Mowrer, O. H., & Jones, H. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.

(Received November 8, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE EFFECT OF INTERTRIAL RESPONDING ON CONDITIONING AND EXTINCTION OF AVOIDANCE BEHAVIOR[1]

J. V. MURPHY[2] AND R. E. MILLER

School of Medicine, University of Pittsburgh

In conditioning experiments which utilize an exteroceptive stimulus as the conditioned stimulus (CS) the conditioned response (CR) is typically defined as a response to the CS before the unconditioned stimulus (UCS) is presented. During the course of conditioning, responses are observed to occur not only in the presence of the CS but also in the intervals between trials. Since the primary interest is typically in CR's, little attention has been directed toward these interval responses (IR's).

Coppock and Mowrer (2) investigated the effect upon avoidance conditioning of permitting interval responses in one group while preventing them in another group. Their results indicated that allowing interval responses significantly facilitated the acquisition of the avoidance response. Unfortunately, as the authors noted in a footnote, procedural details employed to eliminate interval responding in the restricted group may have interfered with conditioning itself thereby making their conclusions somewhat tentative.

In view of the almost complete absence of experimental data in this area an experiment was designed to investigate systematically the effect of interval responding on the course of avoidance conditioning and extinction. More specifically, Exp. I was designed to: (*a*) determine the effect on conditioning and extinction of permitting vs. blocking interval responding; (*b*) determine the effect of blocking interval responses during extinction when interval responding was permitted during the conditioning phase; and (*c*) determine the number of "CR's" and "IR's" which occur when *S*s are subjected to a sham conditioning procedure.

The results of Exp. I suggested the hypothesis that IR's and the high positive correlation typically observed between CR's and IR's during extinction were attributable to the ambiguity which arises from having each half of the apparatus identical in all respects. It was predicted that if the ambiguity were reduced, the rate of CR's and IR's during extinction would be significantly reduced and the occurrence of CR's and IR's would be uncorrelated. Experiment II was designed to investigate this hypothesis.

METHOD

Subjects.—The *S*s in Exp. I were 70 male rats obtained from the Carworth Farms; 38 similar *S*s served in Exp. II. These *S*s were 60–70 days old at the start of the experiments.

Apparatus.—In Exp. I the apparatus consisted of a two-compartment conditioning box with a 2-in. barrier separating the two compartments. The grid floor of each compartment could be electrified independently. This apparatus has previously been described in detail (7, 8). For the present study the apparatus was modified by the insertion of micarda door guides which extended vertically on both sides from the top of the barrier to 10 in. above the apparatus. A Plexiglas door fitted into these guides made it

[1] This investigation was supported by a research grant (M-487C) from the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service.

[2] Now with Systems Development Corporation, Santa Monica, California.

possible to effectively separate the two compartments. A cord attached to the top of the Plexiglas door passed over a rod located above the door guides and was arranged in such a way that a controlled movement of the cord raised the door to the top of the animal compartment. Noise effects were practically eliminated by this arrangement. Two interchangeable Plexiglas doors were constructed. One was a solid door and a rectangular opening was cut in the other door with only a ¼-in. strip on each side projecting into the animal compartment when the door was in the down or closed position. The solid door was used when it was desired to block interval responses and the cut-out door when the experimental condition allowed the occurrence of interval responses.

In Exp. II the apparatus was identical, with the exception that no door or door guides were present. An insert was constructed of ¼-in. plywood which could be fitted into one compartment. This insert was painted black, and when in place, one entire half of the apparatus was black. The remaining compartment was white on the top and all sides.

Procedure.—In Exp. I a door was raised by *E* on all trials with one hand while the buzzer CS was activated with the other. Thus there was no pause between opening the door and presenting the buzzer CS. Similarly, the door was lowered when the buzzer was terminated. Technically, the CS consisted of the opening of the door plus the onset of the buzzer; however, for descriptive purposes the buzzer will be referred to as the CS.

The conditioning trials for each daily test session were preceded by a 1-min. adaptation period followed by the presentation of a buzzer CS. The shock UCS was automatically delivered after 5 sec. if *S* failed to make a barrier-crossing response to the buzzer. The buzzer, or buzzer and shock, were terminated manually when *S* crossed the barrier.

The diminished interval conditioning procedure (**8**) was employed for all *S*s. With this procedure *S*s were given three test periods with an average intertrial interval of 60 sec., three test periods with an average intertrial interval of 50 sec., three test periods with a 40-sec. average, three with a 30-sec. average, and three final conditioning test periods with an average intertrial interval of 20 sec. Extinction was completed with an average intertrial interval of 20 sec. for each test period. Ten trials a day were given for the 15-day conditioning phase and the 15-day extinction period.

The *S*s were randomly assigned to one of four groups with the following experimental conditions: in Group I, intertrial responses were blocked throughout the course of conditioning and extinction; in Group II, intertrial responses were permitted throughout conditioning and extinction; in Group III, the intertrial responses were permitted throughout conditioning and blocked during the extinction phase; in Group IV, the procedure was identical to that of Group II with the exception that these *S*s never at any time experienced the shock UCS. "Intertrial" responding was permitted for this group during the entire 30-day period.

In Exp. II the conditioning and extinction procedures were identical to those described for Groups I, II and III of Exp. I, with the exception that doors and door guides were not involved in any way. The 38 *S*s were randomly assigned to one of the two groups. For Group V, one compartment was black and the other white; for Group VI, both halves of the apparatus were white.

Results

The conditioning and extinction data for Exp. I have been summarized in Table 1. The three conditioning groups (Groups I, II, and III) performed similarly throughout conditioning. The learning curves for the three groups approached the asymptote on Day 8 of conditioning. In addition, the two groups in which interval responses were permitted during conditioning (Groups II and III) did not differ appreciably in the

TABLE 1

The Mean Number of CR's and IR's During Conditioning and Extinction

Group	*N*	Conditioning		Extinction	
		CR's	IR's	CR's	IR's
Experiment I					
I	16	101.19	blocked	87.88	blocked
II	17	107.29	21.06	98.47	43.18
III	18	106.56	23.89	82.78	blocked
IV	18	2.53	66.78	2.81	66.28
Experiment II					
V	15	95.87	22.2	72.93	33.47
VI	17	103.94	18.7	112.06	64.71

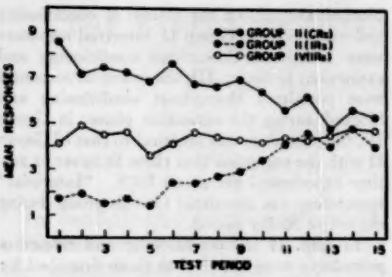

FIG. 1. The mean incidence of responses during extinction in Exp. I.

number of interval responses which occurred (Table 1).

A criterion found to be highly satisfactory in previous studies is to exclude *Ss* failing to respond at or above the 70% level for three consecutive days during conditioning. In the three conditioning groups only one *S* from Group I failed to meet this criterion.

The rate of interval responding for the conditioning groups (Groups II and III) was very similar throughout the conditioning phase. The mean incidence of IR's on the first day was 3.7 with the approximate asymptote of 1.4 attained on the second day. In contrast, the sham conditioned *Ss* had a mean incidence of 9 IR's on the first day with the approximate asymptote of 4.2 approached on the third day of conditioning. Fisher's exact test (4) indicated that *Ss* subjected to the sham conditioning procedure performed significantly more interval responses ($P = .01$) than Groups II and III during the conditioning phase. In addition, the sham conditioned *Ss* (Group IV) performed significantly fewer "CR's" than Groups I, II, and III in both conditioning ($P = .01$) and extinction ($P = .01$). No significant differences in the number of CR's in extinction were found between Groups I, II, and III as tested by Fisher's exact test (4).

In the extinction phase, interval responses were permitted only in Group II and in the sham conditioning group (Group IV). Fisher's exact test indicated that the differences in interval responses for Groups II and IV were not significantly different during this period. The mean number of interval responses throughout extinction for Groups II and IV are depicted in Fig. 1 along with the mean numbers of CR's for Group II.

The conditioning and extinction data for Exp. II also have been summarized in Table 1. The incidence of CR's and IR's for the two groups is shown in Fig. 2. Although no significant differences in either CR's or IR's were found between Groups V and VI in the conditioning phase, these groups diverged markedly in extinction. A repeated measures analysis of variance (3) indicated that *Ss* conditioned and extinguished with black and white compartments (Group V) performed significantly fewer CR's ($P = .01$) and significantly fewer IR's ($P = .04$) in extinction than *Ss* conditioned and extinguished with both compartments white (Group VI).

One aim of Exp. II was to test the prediction that no significant correlation would occur between CR's and IR's when one compartment was

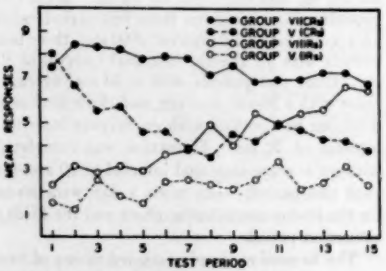

FIG. 2. The mean incidence of conditioned and interval responses in extinction for Exp. II.

white and the other black. The product-moment correlation between CR's and IR's in extinction for *Ss* learning under the all-white condition was .66 ± .08; for the black-white situation $r = .75 \pm .12$. Although a significant relationship was found in both groups, an inspection of the data indicated the presence of a biasing factor. Low responders in terms of CR's are almost invariably low responders in terms of IR's. That is, the correlation typically approaches unity in the case of *Ss* making very few or zero CR's in extinction. Due to this restriction of IR's on the lower end of the scale a differential bias occurs in the two groups because of the significantly lower number of CR's in extinction for *Ss* under the black-white condition. To eliminate this source of bias the correlation between CR's and IR's in extinction was computed only for the top half of each group in terms of CR's. For the black-white compartment *Ss* (Group V) the r between CR's and IR's during extinction was .13 ± .39, which is clearly not significant. For Group VI this correlation was .68 ± .20, which is significant beyond .01.

Discussion

Experiment I.—The present studies were designed in an attempt to assess the importance of the intertrial responding which is typically observed in conditioning studies. As indicated by the data summarized in Table 1, it is apparent that the occurrence or nonoccurrence of interval responding produced no significant alteration in either conditioning or extinction. This evidence is contrary to that reported by Coppock and Mowrer (**2**) who found a significant retardation over the last five days of conditioning when interval responses were prevented. They employed a bar-touching task and intertrial responses were prevented by withdrawing the bar between trials. A trial was initiated by reinserting the bar and presenting the CS. As the authors noted, the appearance of the bar for the experimental group was in an ideal relationship to become a CS for avoidance behavior. Since evidence is available (**9**) which indicates that a CS in avoidance learning acquires fear-evoking qualities, it suggests the possibility that the low rate of response in the response-blocked condition of Coppock and Mowrer might be more parsimoniously attributed to an acquired fear of the bar itself. It appears that this artifact is sufficient to account for the significant differences observed. Unfortunately, Coppock and Mowrer did not obtain extinction data on their rats. It may be noted that in the present experiment the movement of the door is similarly a part of the CS complex. However, this presents no interpretative problem since the moving door was present in both experimental and control groups of the present study.

Although the trend of IR's during conditioning is confounded somewhat in the present study with a progressively diminished intertrial interval, it is important to note that the rate of interval responding had reached an asymptote in all groups before any modification was made in the intertrial interval.

The decreasing incidence of response during the first three test periods for animals subjected to a sham conditioning procedure with no increase in the incidence of "CR's" indicates that the downward trend typically observed in conditioning groups (**5, 6**) may be parsimoniously attributed to a decrement in exploratory behavior plus the inhibitory effect of shock. Although there is a decreasing incidence of response for the sham conditioned animals, it is important to note that the response rate is significantly lower in the two conditioning groups.

Experiment II.—In analyzing the data from Exp. I it was somewhat surprising to find that there was no significant correlation between CR's and IR's during extinction for the animals which were permitted to respond freely during extinction. This observation suggested

the possibility that the problem was altered to some extent by the inclusion of a moving door. This alteration may be attributed either to (*a*) the movement of the door which may have served to direct the attention of the animal to the required location of the response; or (*b*) the door guides may have provided an additional cue that the response had been completed. Since the movement of the cut-out door and the noise involved were very slight, attention was focused on the door guides which served to clearly differentiate the two compartments. If the differentiation of the two compartments is considered to have made the situation less ambiguous for the animal, an enhanced effect might be anticipated if the two compartments were markedly different in color. This line of reasoning suggested the hypothesis that if each compartment were clearly discriminable, the occurrence of IR's would be uncorrelated with the incidence of CR's. In addition, if the information gained on a missed trial in extinction is less ambiguous, it should be utilized more effectively in solving the problem posed during extinction; consequently, it was predicted that *S*s learning in the black-white situation should perform fewer CR's in extinction than *S*s learning in the all-white situation. These predictions were confirmed by the results of Exp. II. When the situational cues were clearly defined by having each compartment a different color (Group V), there was no longer a significant correlation between CR's and IR's in extinction and the number of CR's during extinction was significantly reduced.

One of the most perplexing and paradoxical aspects of avoidance learning has been the failure to find any correlation between the number of CR's in conditioning and the number of CR's in extinction (1). This has been surprising since it seems clear that in the extinction phase one is only measuring behavior which was established in the acquisition phase. With the interpretation that the door guides employed in Exp. I made the problem less ambiguous, it appeared desirable to determine the relationship between CR's in conditioning and extinction for animals permitted to respond freely in Exp. I. The r for these data was $.52 \pm .18$ ($P < .05$).

Experiment II provides a check on this interpretation since for the all-white condition no door guides were present between compartments and the correlation of CR's in conditioning and extinction should be of zero order. With the black-white arrangement the compartments were clearly differentiated and the relationship between acquisition and extinction scores should be highly significant. It may be noted that these correlations were computed on the entire group of rats in each case. The previously described problem of low responders does not bias these data since the possibility of low responders in conditioning is eliminated by the acquisition criterion. The r between CR's in conditioning and extinction for the all-white arrangement was $.20 \pm .15$, which is not significant. A similar r of CR's in acquisition and extinction for the black-white arrangement was $.60 \pm .17$ ($P < .01$). Thus it appears that the usual lack of correlation between acquisition and extinction scores in avoidance learning may be attributable to the ambiguity which such a situation typically contains.

In future studies the use of a black and white compartment shuttle box may be of considerable value when an unusually prolonged extinction period is anticipated. The use of discriminably different compartments would make it possible to extinguish *S*s in a reasonable period of time without reducing the acquisition criterion to such a point that classical and instrumental conditioning are hopelessly confounded.

An additional full scale experiment was completed in an attempt to investigate the possibility that IR's represent an attempt by *S* to avoid the CS for the next trial by utilizing the same response which has been regularly employed to avoid the unconditioned stimulus. The procedure and apparatus were identical to those described for Group VI (all-white compartments) with the exception

that if an interval response occurred the CS and UCS for the next trial were omitted. It was predicted that barrier-crossing responses would increase to a high level and eventually reach a steady state since by performing an IR the CS could be avoided and fear minimized in the situation, i.e., a more complete solution was available. One group ($N = 13$) was conditioned for 15 days and another ($N = 15$) for 30 days. No definitive increase in total responding was observed.

Although we have as yet no answer to the baffling question of the motivation and reinforcement for IR's, the present studies indicate that the occurrence or nonoccurrence of IR's need present no serious question in the interpretation of CR's in this experimental situation.

Summary

Experiment I was designed to assess the effect of intertrial responding on the course of conditioning and extinction of a barrier-crossing avoidance response. Four randomly selected groups of rats were employed. In Group I, interval responses (IR's) were blocked throughout conditioning and extinction. In Group II, IR's were permitted throughout. For Group III the intertrial responses were permitted in conditioning and blocked during extinction. Group IV was subjected to a sham conditioning procedure with no shock UCS present at any time and intertrial responses freely permitted throughout.

The amount of interval responding for the two conditioning groups in which IR's were permitted during conditioning was significantly lower than in the sham conditioning group. The occurrence or nonoccurrence of IR's produced no significant effect on the course of either conditioning or extinction. Conclusions: (*a*) permitting or blocking IR's produces no significant effect on the course of either conditioning or extinction; (*b*) the incidence of IR's is significantly decreased by the shock UCS.

Experiment II was designed to determine the effect of having each compartment discriminably different in color. One group was trained with one black and one white compartment while another group was trained with both compartments white. The use of different colored compartments eliminated the significant correlation between CR's and IR's in extinction which was observed in *Ss* learning under the similar-color compartment situation.

A significant positive correlation was obtained between CR's in conditioning and CR's in extinction for *Ss* learning under the black-white compartment situation. The repeated failure to find such a relationship in avoidance learning was attributed to the ambiguous aspects typically employed for avoidance studies. Conclusions: (*a*) in a two-compartment situation with each half identical a significant correlation occurs between CR's and IR's during extinction; (*b*) with different colored compartments no significant correlation occurs between CR's and IR's during extinction; (*c*) with identical compartments no significant correlation is found between CR's in conditioning and extinction; and (*d*) with different colored compartments the correlation between CR's in conditioning and extinction is significant and positive.

References

1. Brogden, W. J. Acquisition and extinction of a conditioned response in dogs. *J. comp. physiol. Psychol.*, 1949, **42**, 296–302.
2. Coppock, H. W., & Mowrer, O. H. Intertrial responses as "rehearsal": A study of "overt-thinking" in animals. *Amer. J. Psychol.*, 1947, **60**, 608–616.
3. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Fisher, R. A. *Statistical methods for research workers.* New York: Hafner, 1950.
5. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
6. Mowrer, O. H., & Lamoreaux, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58**, 196–213.
7. Murphy, J. V., & Miller, R. E. The effect of adrenocorticotrophic hormone (ACTH) on avoidance conditioning in the rat. *J. comp. physiol. Psychol.*, 1955, **48**, 47–49.
8. Murphy, J. V., & Miller, R. E. Spaced and massed practice with a methodological consideration of avoidance conditioning. *J. exp. Psychol.*, 1956, **52**, 77–81.
9. Murphy, J. V., & Miller, R. E. The manipulation of dominance in monkeys with conditioned fear. *J. abnorm. soc. Psychol.*, 1956, **53**, 244–248.

(Received November 12, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE RECOGNITION, NAMING, AND RECONSTRUCTION OF VISUAL FIGURES AS A FUNCTION OF CONTOUR REDUNDANCY[1]

NANCY S. ANDERSON[2] AND J. ALFRED LEONARD[3]

Ohio State University

A noticeable recent trend in studies of perception has been the renewed attention given to stimulus properties which influence perceptual responses. This applies in general terms to studies in vision (**8**) and more specifically to studies in the field of hearing (**4**). It is interesting to note that considerable progress was made in the latter field once stimuli were considered as members of specifiable ensembles, i.e., once a probabilistic approach had been adopted in the specification of inputs (**11**). The problem of stimulus specification is very much more difficult for visual material, and for this reason the attempt at statistical specification has been made chiefly with nonsense figures (**3**). Specification of visual stimuli in probabilistic terms has led to an examination of the relation between phenomenological Gestalt concepts and those derived from information theory (**1, 2, 12**), and this type of cross-comparison should become increasingly fruitful.

The present experiments were carried out to investigate further a difference in performance noted in a previous study (**7**) using samples of metric figures generated by two different probability rules. Metric figures resemble small histograms and are generated in $n \times m$ matrices, the heights of successive columns being determined by such procedures as reference to tables of random numbers. The rules for selection of column heights may be varied: from a symmetrical $n \times n$ matrix n^n different figures can be generated when no restriction at all is placed on the selection process, and $n!$ when the selection is limited in such a manner that each of the n column heights must appear once in one of the n positions of each single figure. In the paper cited (**7**), the set of figures generated by the first rule was referred to as the *random* sample of figures; that generated under the second as the *constrained* sample of figures.

Applying information measures, constrained figures convey less information than random figures, and the former are redundant with respect to the latter by a specifiable amount. On this count one might have expected the constrained figures to be dealt with more easily than the others. However, figures in a constrained sample are more homogeneous in several respects than are figures in a random sample (e.g., in the former all figures have the same area, the same black-white ratio, and the same number of contour breaks). Thus, "at a descriptive level the constrained figures certainly give the appearance of being more similar" (**7**, p. 9), and

[1] This research was conducted in the Laboratory of Aviation Psychology of the Ohio State University through support from a contract between the Office of Naval Research and the OSU Research Foundation (Contract No. Nonr-495(02)). The authors wish to express their gratitude to P. M. Fitts and D. Bakan, as well as to their principal assistant, Louise Walker.

[2] Now at International Business Machines Research Laboratory, Ossining, New York.

[3] Now returned to Applied Psychology Research Unit, Cambridge, England.

hence, on this count one might have expected constrained figures to be dealt with less easily. In point of fact, it took 40% more time on the average to identify each constrained figure (**7**, p. 6). Since another form of redundancy (the drawing of a mirror-image of the contour) had been found to facilitate performance, Fitts et al. (**7**) interpreted the above finding as demonstrating that the decrement resulting from constrained sampling was due to the particular form of redundancy used rather than that redundancy per se was detrimental.

This finding and the interpretation given to it clearly opened up two distinct lines of inquiry: first, Which forms of redundancy are perceptually useful? and second, Why are some particular forms detrimental? It is the latter question with which we are concerned in the present paper.

It will be noted that frequent use of the words "sample" and "sampling" will be made, mainly with reference to sets of metric figures. Since it is customary to speak of samples of *S*s, but not samples of items or of stimuli, the present usage may be confusing initially. The concept of sampling is, however, a basic one for a probabilistic approach, and hence the retention of the terms with reference to the figures as well as to *S*s is quite deliberate (cf., **3**, p. 453).

Two series of experiments will be presented. In the first, an attempt was made to isolate two stimulus variables concomitant with the constrained sampling procedure. In the second series, we studied the differential effects on performance of samples of figures generated under three different rules of selection and presented to *S*s in three different tasks. In other words, an examination was carried out of the possible interaction between characteristics of a message and *S*'s use of the message.

Experiment I

It is possible to produce certain types of similarities among figures in order to study the effect of a given type of homogeneity on performance in perceptual tasks. The samples of figures generated to achieve specific homogeneities are redundant with respect to those generated randomly without any sampling restriction. The purpose of the first series of experiments was to investigate the effects of homogeneity of area and homogeneity of detail separation.

Method

Stimuli.—Four samples, each containing eight 8×8 metric figures, were used. Two of these were the random and constrained samples described above and used previously by Fitts et al. (**7**). A third sample of figures, Adjusted I, was selected with the restriction that all figures have the same area, equal to the area of the constrained figures. A fourth sample, Adjusted II, was selected with the restriction that in any given figure no two adjacent column heights should be the same.

Task.—The standard identification task was used. This task has been described in detail elsewhere (**7**). In brief, the eight figures of a sample are replicated in each of six rows of a sorting board, the order of figures being random and different in each row. A copy of one of the figures is placed at the top of the board. The task is to identify this figure in each row by turning it over. The same procedure is followed for each figure of the sample. The time required to do this is measured on a Standard Electric timer to the nearest .01 sec.

Subjects.—The 32 male and female *S*s were members of introductory psychology courses at Ohio State University.

Procedure.—Group 1, consisting of 16 *S*s, performed on the Random, the Constrained, and the Adjusted I samples. Group 2, consisting of 16 *S*s, worked with the Random, the Constrained, and the Adjusted II samples. In each group all *S*s were tested with all three samples of figures, the order of presentation being independently randomized for each *S*. The experiment was given in two 1-hr. sessions on two suc-

cessive days, each *S* participating in each condition twice on each day, a total of 12 trials for the two days. A trial consisted of eight items. An item consisted of the identification of a designated figure in each of the six rows of the sorting board.

The *S*s were told to go as fast as they could, but to correct any errors in any one row before going on to the next. On completion of a trial, *S*s were told the summed time for all eight items.

Scoring.—The sum of the times for the two trials under the same condition on Day 2 was divided by 96 (2 trials × 8 items × 6 rows) to give a mean time per figure per row, for each sample and for each *S*. Only the performance data for the second day will be presented.

Results

For Group I the mean times per figure per row (in seconds) were 2.00 for the Random, 2.54 for the Adjusted I, and 3.36 for the Constrained figures. For Group 2 the mean times were 2.18 for the Random, 2.92 for the Adjusted II, and 3.52 for the Constrained figures. In both groups, analysis of variance indicates significant effects due to the rules of sampling of column heights ($P < .01$). As in previous studies, *S*s required longer identification times for constrained than for random samples of figures, and the adjusted samples were of intermediate difficulty.

The Mann-Whitney *U* test (**10**) of significance between groups using the two samples of adjusted figures showed that mean times for Adjusted II were significantly greater than for Adjusted I ($P < .01$). As a check on the comparability of the two different groups who used the same samples of random and constrained figures, it should be mentioned that there were no significant differences for scores on random figures or for scores on constrained figures.

The results of this experiment were consistent with those of the previous study (**7**) and indicated that all of the forms of constraint used when selecting column heights in the generation of metric figures resulted in a performance decrement, and that some constraints had a greater effect than others. The implications of these findings will be referred to in the final discussion section.

There were now a number of further possibilities for research. Two of these were the investigation of further variants of sampling constraints and the interaction between different types of constraints and the learning and performance requirements of different tasks. It was decided to follow the second of these alternatives.

Experiment II

In the second series of studies we decided to make use of figures generated by three different rules of sampling and to use samples of 12 rather than of 8 figures. The sizes of the Random, Constrained, and Adjusted II samples used in Exp. I were therefore increased accordingly. We used the Adjusted II in preference to Adjusted I sample because the type of constraint employed there was more closely related to that employed in the standard constrained sample, i.e., the constraint affected column interdependency rather than any property of the whole figure. The three tasks chosen for study were a classification task, a paired-associate learning task, and a reconstruction task. The first was a task rather similar in most respects to the standard identification task used earlier; the second required a verbal rather than a motor response; and the third was one in which *S*s had to reconstruct the details of each figure on a physical model. In Exp. II a spatial perception pretest (**5**) was to enable the population from which *S*s had been sampled to be specified. These test results were also used for matching purposes.

Method

Tasks and apparatus.—The three tasks were as follows: (*a*) In the classification task, *S*s were presented with stacks of 72 stimulus cards (6 replications of 12 figures) arranged in random order. The *S*s were required to place each figure in turn into one of 12 boxes, below each of which was a copy of one of the 12 figures. The score was the time required to sort once through one set of 72 figures, as read from a Standard Electric timer. (*b*) In the paired-associate task, stimuli were reproduced on 2 × 2-in. slides and projected on a screen in one of two ways: for practice sessions, stimuli and letters were projected together; for test sessions, stimuli were projected by themselves. Stimuli were presented at the rate of one every 5 sec., with an exposure time of 3 sec. Rate of presentation and duration of exposure time were controlled by a disc rotating in front of the projector. The figure image on the screen was 18 × 18 in. (*c*) In the reconstruction task, *S*s were presented with one figure which was exposed for .5 sec. at a time. They were required to reconstruct this figure using a physical model on which they could adjust eight columns to the correct heights. Each stimulus figure was exposed repeatedly until *S* succeeded in reconstructing it correctly. Stimuli were exposed in the frontal horizontal plane by a shutter arrangement, the action of which was under *S*'s control. The duration of exposure was uniform and not under *S*'s control.

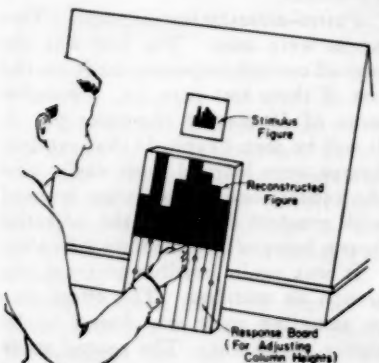

FIG. 1. Apparatus used in the figure reconstruction task. The *S* exposed the stimulus (upper) figure for .5 sec., then moved the individual bars of the lower figure to form the same contour, using as many exposures as necessary to complete the task.

The general appearance of the apparatus is shown in Fig. 1.

Subjects.—The 144 male and female *S*s used in Exp. II were members of psychology courses at Ohio State University.

Procedure.—The over-all design is best visualized in a 3 × 3 block diagram having the three samples of figures along one dimension and the three tasks along the other. In each of the nine cells 16 different *S*s were used. These 16 *S*s were selected from a larger population on the basis of the pretest, matching means and variances for all nine groups of *S*s. Within this over-all design, procedures varied depending on the task used. For the classification task each *S* was required to deal with his particular sample twice in succession. For the paired-associate task, practice and test sessions alternated: there were four practice sessions during which two lists of 12 items were presented, each practice session being followed by a test session of one list, except for the final practice session which was followed by three lists. The orders of presentation for the 12 figures differed within each practice and test list. For the reconstruction task, *S*s were presented with repeated exposures of one of the 12 figures until they had reconstructed it correctly. Then a new figure was presented with repeated exposures. This was continued until each *S* had reconstructed all 12 figures correctly. A different order of presentation was used with each *S*.

All *S*s were given the California Survey of Space Relations Ability (5) as a pretest a few days before their first experimental session. Pretesting required about 20 min., while 1 hr. was allowed for the experimental sessions. For the classification and reconstruction tasks *S*s were tested singly, but for the paired-associate learning task the experiment was conducted with groups of 16 *S*s. Instructions can be paraphrased as follows: (*a*) For the classification task, *S*s were instructed to work as fast and as accurately as possible. Very few errors were made and these were not taken into account in scoring. (*b*) For the paired-associate learning task, *S*s were provided with separate booklets for the practice and test sessions. They were instructed to write down their responses on a different sheet of the booklet for each stimulus. During practice they were encouraged to note special features of the figures which might help in learning. The *E* read out the correct order of responses after each test session, and after *S*s had put down their pencils. (*c*) For the reconstruction task, *S*s were instructed to make successive adjustments to the slides of the physical model after each exposure of the test figure. Having moved all eight column heights once, i.e.

having achieved a "first approximation" of the figure, they were informed of those columns which were in error, but not the amount and direction of the error. They were then permitted to take additional half-second looks at the figure and make further adjustments until they were satisfied.

TABLE 1

SUMMARY OF THE RESULTS OF EXP. II

Task	Sample of Figures		
	Random	Adjusted	Constrained
Classification:			
Time to sort six replications of each of 12 figures			
Mean	4.92	6.37	7.19
SD	1.09	1.19	1.08
Paired-Associate Learning:			
Number of items correct on last test trial, three lists of 12 figures each			
Mean	30.06	20.56	15.31
SD	5.87	6.93	7.20
Correct item score—consistency score on last three test lists			
Mean	8.44	4.56	2.38
SD	8.12	7.25	4.86
Reconstruction:			
Score 1—number of exposures to reproduce 12 figures correctly			
Mean	98.37	101.81	82.06
SD	21.65	21.29	28.35
Score 2—number of exposures to complete first approximation of 12 figures			
Mean	63.87	63.37	50.56
SD	14.66	14.78	16.54
Score 3—average number of column heights manipulated per exposure for 12 figures			
Mean	23.90	23.07	30.63
SD	5.64	5.50	17.03
Score 4—column heights in error at first approximation for 12 figures			
Mean	41.00	41.94	47.68
SD	13.86	22.19	27.99

Results

The mean scores on the pretest for the nine groups of 16 *S*s each, ranged from 64.00 to 64.62, with corresponding *SD*'s ranging from 17.11 to 18.39. Table 1 shows the means and *SD*'s for all scores on all the experimental tasks. Table 2 shows the summary of tests of significance for the experimental effects.

Classification task.—It will be seen (Table 1) that the least time was required to classify the random sample and the most time to classify the constrained sample, the difference being about 46%. The results of an analysis of variance performed on these data are summarized in Table 2. Matching having been found to be successful, the appropriate test of significance was used. The effect due to differences in the degree of restriction imposed by the sampling procedures for stimulus figures was highly significant.

Paired-associate learning task.—Two scores were used. The first was the sum of correct responses made on the last of three test lists, i.e., a possible score of 36 correct responses per *S*. It will be seen (Table 1) that random figures were learned most easily and the constrained figures were learned with greatest difficulty, the adjusted figures being of intermediate difficulty.

It was again possible to treat the groups as matched. The effect due to sampling rules was found to be highly significant. The second score used is a form of criterion score and referred to as the "correct item score." For each *S* the number of figures correctly identified through the three final test lists was counted (see Table 1). This score appears to be somewhat more sensitive than the

TABLE 2

SUMMARY OF TESTS OF SIGNIFICANCE FOR THE DATA OF EXP. II

Task and Test of Significance	Source	*df*	Mean Square or Standard Normal Deviate	*P*
Classification (Analysis of Variance)	Sampling Rules	2	21.15	<.01
	Subjects	15		
	Error	30	.97	
	Total	47		
Paired-Associate Learning Number Correct (Analysis of Variance)	Sampling Rules	2	894.33	<.01
	Subjects	15		
	Error	30	30.02	
	Total	47		
Consistency Score (Mann-Whitney *U*)	Random > Constrained		−3.00	<.01
	Adjusted > Constrained		−2.35	<.01
	Random > Adjusted		−3.00	<.01
Reconstruction Score 1 (Analysis of Variance)	Sampling Rules	2	1781.30	NS
	Error	45	614.20	
	Total	47		
Score 2 (Analysis of Variance)	Sampling Rules	2	911.10	<.05
	Error	45	251.50	
	Total	47		
Score 3 (Mann-Whitney *U*)	Constrained > Random		−1.39	.08
	Constrained > Adjusted		−1.92	.03
	Random > Adjusted		−.37	.35
Score 4 (Mann-Whitney *U*)	Constrained > Random		−.40	.37
	Constrained > Adjusted		−.85	.20
	Adjusted > Random		−.43	.34

total correct score shown above: on the latter score, constrained figures obtained 49% fewer correct responses than random; with the correct item score, the percentage difference was 72% (results of Mann-Whitney's *U* Test, $P < .01$).

Reconstruction task.—A number of different scores were used to evaluate performance on the reconstruction task. This task was the only one of the three in which prematching did not have a significant effect. Accordingly, a random assignment design was used for the analyses of the data. The first score examined was the number of exposures required to reproduce each figure perfectly. The analysis of variance using this score showed that there was no significant effect due to rules of sampling.

The second score taken was the number of exposures required to complete "the first approximation," i.e., the number of exposures required to manipulate each column at least once. It will be seen that there was little difference using this score between the random and adjusted figures, but that the constrained figures required fewer exposures than either of the other two. The effect due to rules of sampling was significant at the .05 level.

The third score taken was the number of columns moved per exposure (Table 1). There was a close similarity, again, between random and adjusted; however, the constrained figures showed a larger number of columns moved than the other two sets of figures. In view of the considerable difference in the type of distributions, the effects were tested by the Mann-Whitney *U* test, which indicated a significant difference between the constrained and the adjusted samples (Table 2).

The fourth score was the amount of error found at the first approximation. This score was obtained by taking the sum of the absolute differences between actual and observed column heights for each column and for each figure. It can be seen that there were no significant differences due to sampling rules reflected in this score (Mann-Whitney *U* Test).

Generally speaking, it appears that it was easier to achieve a first approximation of constrained figures, but that this initial advantage was lost when complete accuracy of reconstruction was required. A further point to note is that in the reconstruction task the adjusted sample throughout had scores more similar to the scores of the random than to the constrained sample.

Discussion

The major purpose of the two series of experiments was to obtain an answer to the question: Why is a particular form of redundancy detrimental to performance in an identification task? In Exp. I, the two different methods of constraining the sampling of column heights had a detrimental effect on performance in the identification task. However, restricting the relation of adjacent figure details was more detrimental than restricting the over-all area of the figure.

In Exp. II, performance measures on two new tasks, that of classification and paired-associate learning, showed the same relation among the sets of figures generated as had the original identification task. The agreement between identification and classification may be accounted for by the fact that *S*s were required to use similar perceptual processes—a matching of figures. This is consistent with results by Leonard, Clarke, and Staats (9) which showed high positive transfer from a classification to the standard identification task. The agreement between identification, classification, and paired-associate learning is consistent with the results of Adams et al. (1), who found a high positive correlation between paired-associate and identification measures. This relation is of theoretical importance in that it emphasizes the role of identification, discrimination, and classification in learning to name samples of objects as stressed by Adams et al.

The reconstruction task, however, shows a different relation among the three samples of figures than the other tasks. All measures in this task indicate a close similarity between random and adjusted samples, whereas in other tasks, measures indicated that the adjusted sample was more closely related to the constrained sample. It should be noted that the reconstruction task differs from the other tasks in that, in the former, it is necessary to note every detail of a figure, and each figure was presented in isolation—that is, not in context with the sample.

These differences between the two sets of tasks and, more specifically, the different measures obtained in the reconstruction task provide a number of clues about possible effects on performance of different rules of sampling figure details. Note that it took significantly fewer exposures to achieve a first approximation of the constrained figures than for the figures in the other samples. Also, a greater number of columns were manipulated per exposure on the constrained figures. But, there is no differ-

ence between the number of exposures required to complete a perfect reconstruction for any of the three samples. This indicates that it may be easier to perceive the general outline of a constrained figure, but perfect reconstruction is as hard to achieve for these as for the other two samples of less restricted figures. It is concluded that the advantage of random figures in a recognition task must result from *S*'s ability to make his discriminations on the basis of distinctive parts of each figure. It is therefore assumed that randomness provides more and/or easier to discriminate combinations of details and thus permits *S* to reduce the amount of information he must process before he can respond. Thus, although redundancy reduces total figure information, in certain cases it may greatly increase the information that *S* must process before he can identify a particular figure. In line with this analysis were the findings from the first experiment, in that a restriction on sampling of adjacent details had a more detrimental effect on speed of identification than a restriction which applied to over-all area.

In general, the present findings support the original interpretation of Fitts et al. (7) and explain it even further. The constraints employed introduce redundancy in an unsatisfactory manner because these constraints reduce the probability of distinctive cues appearing in each of a given sample of metric figures. One suspects that constrained figures would be judged by most observers to be phenomenologically "better" than the random ones. Yet their use in several tasks has been shown to have a definitely detrimental effect on performance. One is reminded of the finding that aesthetically good sets of letters gave consistently poorer performance than aesthetically bad sets in legibility studies (6). The possibility of cross-comparisons between phenomenologically derived concepts and those based on information theory was mentioned in the introduction. More generally, one may now suggest that with increasing ability to specify stimulus properties in probabilistic terms the time seems ripe for either an extension of information measures or a new measure to predict why different methods of introducing redundancy result in differential performance in perceptual tasks.

Summary

The present study extends and analyzes the previously reported finding that one form of redundant metric figures (made redundant by the sampling restriction that a particular contour detail could occur only once) was more difficult to identify among a sample of similar figures than were random (nonredundant) metric figures. In Exp. I, two intermediate levels of redundancy (restricting total figure area and restricting the relation of adjacent details) were compared with the more extreme levels studied earlier. In Exp. II, the perceptual task was varied so as to require (*a*) classification of a series of figures, (*b*) paired-associate learning of figure names, and (*c*) complete reconstruction of figures.

Use of the two intermediate levels of redundancy produced figures that were more difficult to identify than were random figures, but less difficult than the original redundant (constrained) figures. Homogeneity among figures which was produced by restricting the random relation of adjacent contour details (no two adjacent columns the same height) had a greater effect than did homogeneity produced by holding over-all figure area constant.

Two of the new perceptual tasks studied in Exp. II showed effects of levels of redundancy which were similar in direction and relative magnitude to the effects found with the standard identification task. These were the classification task and the learning task. Levels of redundancy had minor over-all effects on the efficiency of complete figure reconstruction, however.

These results are interpreted as supporting the hypothesis that *S*s will find ways of utilizing only the amount of information called for by a perceptual task. Thus, in identification, classification, and learning tasks they will use only distinctive details of a figure. Such distinctive details are more likely with the use of random (nonredundant) figures. Only in the reconstruction task is it necessary to respond to all of the information (all details) in a figure. The results suggest certain interesting differences between the data obtained for phenomenological and performance criteria, and between studies of the properties of individual figures and the properties of samples of figures.

References

1. Adams, O. S., Fitts, P. M., Rappaport, M., & Weinstein, M. Relations among some measures of pattern discriminability. *J. exp. Psychol.*, 1954, **48**, 81–88.
2. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–194.
3. Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.
4. Broadbent, D. E. Growing points in multichannel communication. *J. acoust. soc. Amer.*, 1956, **28**, 533–535.
5. Case, H. W., & Ruch, F. Survey of space relations ability. Los Angeles: California Test Bureau.
6. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.
7. Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51**, 1–11.
8. Gibson, J. J. *The perception of the visual world*. Boston: Houghton, Mifflin, 1950.
9. Leonard, J. A., Clarke, H. W., & Staats, S. R. Factors contributing to general versus specific perceptual learning. *J. exp. Psychol.*, 1957, **53**, 324–329.
10. Mann, H. B., & Whitney, D. R. The *U* Test. *Ann. math. Statist.*, 1947, **18**, 50–60.
11. Miller, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
12. Rappaport, M. The role of redundancy in the discrimination of visual forms. *J. exp. Psychol.*, 1957, **53**, 3–10.

(Received November 13, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

BODY TEMPERATURE AND PSYCHOLOGICAL RATINGS DURING SLEEP DEPRIVATION[1]

EDWARD J. MURRAY,[2] HAROLD L. WILLIAMS, AND ARDIE LUBIN

Walter Reed Army Institute of Research

Kleitman (1) has stated that body temperature is closely associated with many of the phenomena of sleep and wakefulness, although he thinks that body temperature, itself, is simply a convenient way of measuring muscle tonus. Kleitman (1), and Kleitman and Jackson (2) have shown that body temperature has a diurnal variation which persists during sleep deprivation and which is positively related to the efficiency of performance on such tasks as reaction time and color naming. Kleitman (1) goes on to say that body temperature is negatively related to subjective feelings of sleepiness but presents no evidence for this assertion other than an inference from the performance data. The purpose of this study was to test this hypothesis by comparing psychological ratings of sleepiness and fatigue by *Ss* and *Os* with body temperature measures during 98 hr. of sleep deprivation.

METHOD

Subjects and observers.—The *Ss* were 15 enlisted men who had volunteered for a study on the effects of sleep loss on a number of psychological and physiological variables. The group of *Os* was composed of 15 control *Ss*, three noncommissioned officers, four Army nurses, and a social science assistant. The experimental *Ss* were studied in groups of five on a comfortably heated ward which is more fully described elsewhere (4). After a control period, *Ss* were kept awake by gentle methods from 6:00 A.M. on a Friday morning to 8:00 A.M. on the following Tuesday morning for a total of 98 hr.

Rating scales.—Two four-point rating scales were developed on the basis of preliminary studies. The *sleepiness* scale consisted of the following steps: 1. wide awake; 2. getting sleepy; 3. pretty sleepy; and 4. trouble staying awake. The *fatigue* scale consisted of the following steps: 1. not tired; 2. little tired; 3. pretty tired; and 4. dead tired. Each step in both scales was printed on an index card handy for sorting.

Self-rating procedure.—During the 98 hr. of sleep deprivation each *S*'s oral temperature was measured 10 times during each 24-hr. period at: 9:00 A.M. and 12:00 M.; 3:00, 6:00, 8:00, 10:00, and 12:00 P.M.; and 2:00, 4:00, and 6:00 A.M. An extra reading at 8:00 A.M. on the last day brought the total number of readings to 41. While the thermometer was in *S*'s mouth, he was seated at a table and asked to make several ratings. He selected the one statement from the sleepiness scale and the one statement from the fatigue scale which best described his subjective state at that particular moment. Then, using the sleepiness scale, he was asked to describe each of his four fellow *Ss* in the experimental group.

Observer rating procedure.—Shifts were arranged so that one nurse and one noncommissioned officer were always on duty during the 98 hr. of sleep deprivation. Immediately after the temperatures were taken, the nurse and the noncommissioned officer independently selected a statement from the sleepiness scale which best described each of the experimental *Ss*. In addition to the nurse and noncommissioned officer ratings, the five members of each control group rated the experimentals. They made six ratings a day between 9:00 A.M. and 10:00 P.M., but, of course, had to sleep during the night. A social science assistant continued the sleepiness ratings from 12:00 P.M. to 6:00 A.M. each day. The nurse rating, the noncommissioned officer rating, and the combined control group and social science assistant rating over the 98 hr. of sleep deprivation showed an average multiple correlation of .80. Therefore, these three sets of ratings were averaged to obtain an over-all observer rating for each experimental *S*.

RESULTS AND DISCUSSION

The average body temperature during the sleep deprivation period is

[1] The authors wish to thank all of the staff members who helped with the ratings and Melvin Cohen who helped with the statistical computations.

[2] Now at Syracuse University.

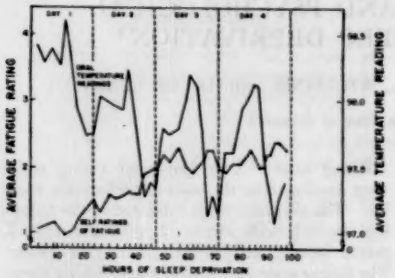

FIG. 1. Average oral temperature measures in degrees Fahrenheit and average self-ratings of fatigue for 15 *Ss* during 98 hr. of sleep deprivation.

shown in Fig. 1 along with the average self-rating of fatigue for comparison. It can be seen that body temperature shows a persisting diurnal variation during sleep loss as Kleitman reports. However, body temperature also shows an over-all decrease while self-ratings of fatigue show an increase with hours of sleep deprivation.

Table 1 shows the intercorrelations between each of the four ratings, body temperature, and hours of sleep deprivation. All four psychological ratings have high and positive correlations with hours of sleep deprivation and with each other. On the other hand, body temperature is negatively correlated with each of the four psychological ratings as well as with hours of sleep loss. However, before concluding that sleepiness ratings in general are inversely related to body temperature, it is necessary to partial out the effects of hours of sleep loss with which both ratings and temperature are correlated. When amount of sleep loss is held constant, the partial correlations with body temperature are: observer rating of sleepiness, −.40; self-rating of sleepiness, −.46; fellow *Ss* rating of sleepiness, −.47; and self-rating of fatigue, −.76. Therefore, sleepiness ratings in general appear to be inversely related to body temperature even when hours of sleep deprivation are held constant. These results strongly support Kleitman's hypothesis about body temperature and sleepiness but, of course, have no bearing on the relationship between body temperature and muscle tonus.

The average sleepiness ratings by observers, fellow *Ss*, and the *Ss* themselves during the course of the 98 hr. of sleep deprivation are shown in Fig. 2. It can be seen that *Ss* rate themselves as less sleepy than they do their fellow *Ss*. Furthermore, both *Ss*' self-ratings and their ratings

TABLE 1

PRODUCT-MOMENT *r*'S BETWEEN FOUR RATINGS, BODY TEMPERATURE, AND HOURS OF SLEEP DEPRIVATION

	Fatigue-Self	Sleep-Self	Sleep-Fellows	Hours of Sleep Loss	Body Temp.
Sleep-Obs.	.94	.73	.92	.94	−.66
Fatigue-Self		.78	.91	.99	−.67
Sleep-Self			.89	.78	−.69
Sleep-Fellows				.92	−.69
Hours of Sleep Loss					−.59
Body Temperature					

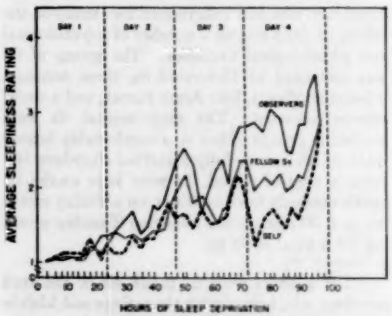

FIG. 2. Average sleepiness ratings by observers, fellow *Ss*, and the *Ss* themselves during 98 hr. of sleep deprivation.

of their fellow *Ss* are considerably below the ratings of the observers. These three differences increase significantly ($P < .001$) during the four days of sleep loss when evaluated by a nonparametric test based on average Spearman correlations of ranked daily discrepancies between pairs of ratings and ranked days of sleep loss (**3**). By the fourth day of sleep loss the differences are significant ($P < .01$) by a sign test. However, right at the end of the sleep deprivation period *Ss* suddenly rate themselves about as sleepy as the observers do. It would appear that up until the last few hours *Ss* denied being sleepy, presumably as an adjustment mechanism for maintaining wakefulness. Conceivably, this then generalized to fellow *Ss*. This interpretation is supported by the fact that self-ratings of sleepiness have the lowest correlations with other ratings and with hours of sleep deprivation.

Summary

Body temperature measures as well as ratings of sleepiness and fatigue were obtained from 15 *Ss* during 98 hr. of sleep deprivation. Observers also rated the *Ss*. Body temperature showed a persisting diurnal variation but also showed an over-all decrease with hours of sleep deprivation. Self-ratings of sleepiness and fatigue, and sleepiness ratings by fellow *Ss* and observers, were positively intercorrelated as well as positively correlated with increasing hours of sleep loss. The four ratings and body temperature were inversely correlated even when hours of sleep deprivation were partialed out. The *Ss* tended to rate themselves as less sleepy than the observers did, which was interpreted as an adjustment mechanism. It was concluded that the results support Kleitman's hypothesis that reports of sleepiness are inversely related to body temperature.

References

1. Kleitman, N. *Sleep and wakefulness.* Chicago: Univer. of Chicago Press, 1939.
2. Kleitman, N. & Jackson, D. P. Body temperature and performance under different routines. *J. appl. Physiol.*, 1950, **3**, 309–328.
3. Lyerly, S. B. The average Spearman rank correlation coefficient. *Psychometrika*, 1952, **17**, 421–428.
4. Murray, E. J., Schein, E. H., Erikson, K. T., Hill, W. F., & Cohen, M. The effects of sleep deprivation on social behavior. *J. soc. Psychol.*, in press.

(Received November 14, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

DEPRIVATION AND GENERALIZATION

W. O. JENKINS, G. R. PASCAL, AND R. W. WALKER, JR.

University of Tennessee

The purpose of this study is to investigate the relationship between drive level and the slope of the generalization function. Miller (4) has suggested that an increase in drive will raise the over-all height of the generalization gradient. Both Brown (1) and Rosenbaum (5) have reported a tendency for the generalization function to be flattened under conditions of high drive. Brown's study was concerned with spatial generalization, and Rosenbaum's with temporal generalization. Both used rats. The present study used pigeons, size generalization, a wide range of stimuli, and extremes of drive level.

METHOD

Two experiments were conducted with food-deprived pigeons as *Ss*. Drive was defined in percentage of satiated body weight. In Exp. I, six birds were divided into a high-drive group of three *Ss* at 70% and a low-drive group of three *Ss* at 80% of body weight. All six birds were subjected to some preconditioning treatments for other purposes (6). In Exp. II, eight birds (not given preconditioning treatment) were thoroughly stabilized at 70% of body weight and another eight at 90%. In both investigations, the birds were conditioned to perform the pecking response in the standard Skinner box for pigeons. After approximately 100 reinforced responses on a 100% basis, they were placed on a 12-min. aperiodic reinforcement (APR) schedule. Training continued until responding stabilized. This procedure required about two to three weeks with a 48-min. session per day. For more details, the reader is referred to a previous paper by Brush, Bush, Jenkins, John, and Whiting (2).

During the conditioning and APR performance, an illuminated spot, 1.4 cm. in diameter, was projected on the pecking window. The diameter of this spot was varied in the generalization tests. Six generalization stimuli were employed with diameters of .2, .6, 1.0, 1.8, 2.2, and 2.6 cm. These were coded, in order, −3, −2, −1, +1, +2, and +3, with the standard being designated 0. The seven stimuli (including the standard) were presented in a nonsystematic order for generalization testing in the following way. If a given *S*'s performance in the first 12 min. of a session was close to his stabilized APR level, the food mechanism was disconnected and one of the seven stimuli was presented for a 12-min. extinction test. The standard training stimulus and the reinforcing stimulus were then reintroduced until responding stabilized after which another generalization test was conducted. Usually *Ss* were subjected to one test a day. Not more than two tests were conducted in any one day. The behavior of *Ss* in Exp. I was erratic, presumably due to their preconditioning treatment, and only one series of generalization tests could be obtained. Two series were completed in Exp. II where behavior was quite consistent except for one *S* in the low-drive group whose performace was so variable its record had to be discarded.

To cut down individual differences and to provide a meaningful figure for assessing the interaction of drive and generalization, a percentage was used. It consisted of dividing responses during the 12-min. test by the number in the preceding 12-min. APR performance interval and multiplying the resultant by 100. Thus if *S* responded 100 times in the first 12-min. APR interval and produced 90 reactions in the following 12-min. generalization test, its score was 90%. This transformation was used for results with all seven stimuli including the standard.

RESULTS

The data are summarized in Tables 1 and 2, and Fig. 1. The tables present the *S*-by-*S* results for the two experiments. Because performance was so variable in Exp. I, the results for the four stimuli most similar to the standard have been lumped as have those for the two most remote stimuli. The findings are strikingly clear and consistent. Large generalization decrement effects occurred for all six *Ss*. The focal effect is the drive differential. As dissimilarity increases,

TABLE 1

GENERALIZATION IN EXP. I AS A FUNCTION OF DRIVE LEVEL IN PERCENTAGE TERMS

Drive	S No.	CS	Near Stimuli	Far Stimuli
Low	28	125	75	16
	29	96	36	14
	30	94	42	0
	Mean	105	51	10
High	25	87	55	36
	26	102	82	72
	27	102	60	49
	Mean	97	66	52

generalization is much greater for the high-drive condition than the low-drive condition. Where the stimuli are most dissimilar, the difference amounts to a factor of five with no overlap in the two sets of three figures.

Analytical statistics are hardly required in such data, but when applied yield high significance. For example, the F values associated with spot size, drive, and their interaction in Table 1 produce probabilities well beyond the .01 level.

TABLE 2

GENERALIZATION IN EXP. II AS A FUNCTION OF DRIVE LEVEL IN PERCENTAGE TERMS

Drive	S No.	Spot Sizes						
		−3	−2	−1	0 (CS)	+1	+2	+3
Low	1	01	34	47	106	55	43	20
	3	01	52	16	109	32	20	10
	4	18	26	67	114	51	65	39
	5	01	08	53	85	53	21	55
	6	02	24	36	72	46	21	16
	7	23	46	75	104	53	48	51
	8	54	38	98	115	58	41	32
	Mean	15	32	56	100	49	37	32
High	9	23	85	103	90	56	57	47
	10	12	80	108	107	80	28	33
	11	55	66	113	110	89	69	74
	12	48	80	66	111	82	51	74
	13	41	80	79	104	79	69	64
	14	13	21	34	90	79	34	19
	15	44	71	104	112	117	85	80
	16	53	99	91	105	83	80	71
	Mean	36	73	87	104	83	59	58

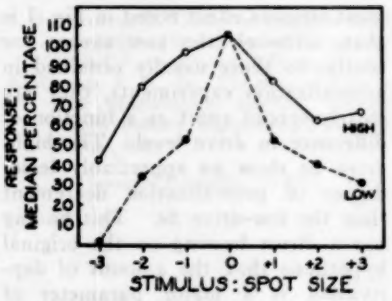

FIG. 1. The relationship between generalization and deprivation in Exp. II.

The results in Exp. II are very similar. As stimulus dissimilarity increases, behavior strength declines systematically. As a matter of fact there are only 16 reversals of the 90 possible in Table 2 and a majority of them are less than 5%. The drive influence is marked, particularly on the most dissimilar stimuli (±2 and ±3) where the distributions come very close to nonoverlapping.[1] Again, standard statistics support inspection by demonstrating the significant effect of drive, stimulus change, and their interaction. A slope constant was computed for each S and a t test was applied across drive levels. The result was significant at the .003 level.

Figure 1 shows, graphically, the results obtained in Exp. II. The

[1] One side matter is pointed up in Table 2 and Fig. 1. Responding was consistently and appreciably greater to the largest generalized stimulus than to the smallest one. The difference in percentage units is about 20 on the average and only 1 of the 15 Ss produced a reversal. This finding may be related to Hull's concept of "stimulus dynamism" (3), but it should be kept in mind that pigeons are "visual" organisms and become immobile in the dark. The smallest spot size (−3) produced barely discriminable illumination to the human eye.

most obvious effect noted in Fig. 1 is that, although the two curves are similar to those usually obtained in generalization experiments, they are clearly spread apart as a function of difference in drive level. The high-drive *S*s show an appreciably lesser degree of generalization decrement than the low-drive *S*s. This finding has a direct bearing on the original hypothesis that the amount of deprivation is a major parameter of generalization; the greater the deprivation, the greater the generalization.

Discussion

In an attempt to arrive at an hypothesis which might account for the results depicted in Fig. 1, the raw data of Exp. II were inspected. They showed practically no overlap for the two drive conditions with no stimulus change. The average number of responses per 12 min. during APR performance was 147 for the low-drive group and 403 for the high-drive condition, indicating, as expected, a greater response strength for the high-drive *S*s. Drive seems to increase generalization by way of increasing response strength.

To test this hypothesis, the three highest and the three lowest APR performers prior to generalization testing were selected from each drive group of Exp. II. An index of generalization decrement was obtained by taking the difference in percentage performance during generalization between the standard training stimulus and the mean of the two most dissimilar stimuli. This method provides a crude but efficient estimate of the slope constant for each *S*. For the high-drive *S*s the difference scores for the high performers were .50, .52, and .43; for the low performers of this condition they were .55, .79, and .74. The two sets of three events do not overlap and a *t* test yields a probability near the .01 level. For the low-drive group, the high responders produced values of .86, .57, and .71 while the low reactors gave .96, 1.03, and .67. The probability against the hypothesis of a common population for these figures lies near the .10 level.

It is clear, therefore, that greater response strength operates to flatten the generalization curve with constant drive level. The results show that drive level *is* a powerful parameter of generalization, but that *within* drive level, response strength is the determining factor which accounts for the flattening effect on the generalization function. The general hypothesis emerges that variables, such as drive, which affect response strength have a corresponding influence on generalization.

These experiments have demonstrated that the slope of the generalization gradient varies with drive level. Generality has been added to the findings by using different drive levels in the two investigations. When these findings are added to those of Brown (1) and Rosenbaum (5), the principle seems well established that deprivation has a marked effect on generalization.

Summary

Two experiments were conducted on the effect of deprivation on generalization. Two groups of pigeons, one at high drive and one at low, were conditioned to a stable response state under aperiodic reinforcement conditions. Generalization tests were then conducted in which the size of the illuminated spot on the pecking window was varied. A direct relationship was found between drive and the slope of the generalization function.

References

1. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
2. Brush, R. R., Bush, R. R., Jenkins, W. O., John, W. F., and Whiting, J. W. M. Stimulus generalization after extinction and punishment: An experimental study of displacement. *J. abnorm. soc. Psychol.*, 1952, **47**, 633–640.

3. HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
4. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
5. ROSENBAUM, G. Temporal gradients of response strength with two levels of motivation. *J. exp. Psychol.*, 1951, **41**, 261–267.
6. WALKER, R. W., JR. The effect of drive on the generalization function. Unpublished doctor's dissertation, Univer. of Tenn., 1956.

(Received November 18, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

GRADIENTS IN RESPONSE PERCENTAGES AS INDICES OF NONSPATIAL GENERALIZATION [1]

BETTINA BASS [2]

State University of Iowa

Stimulus generalization, the elicitation of a response by cues resembling the conditioned stimulus, has been observed in a variety of experimental situations. The paradigm of classical conditioning has been used by some investigators (e.g., **4, 6, 9**) in studies of this phenomenon, whereas others (**5, 8**) have investigated the generalization of instrumentally conditioned responses. Instances of the transfer of voluntary responses have also been reported. For example, the generalization of a voluntary verbal response along a tactual dimension has been described by Gibson (**3**), while Brown, Bilodeau, and Baron (**1**) have observed generalized voluntary manual responses to visual stimuli.

More recently, Brown, Clarke, and Stein (**2**) have found indications of generalization in a situation which involved verbal responses to visual cues. Their *S*s were told that the experiment concerned betting accuracy during a series of simulated horse races. Seven identical light bulbs, spaced at equal intervals in a semicircle in front of *S*, represented seven horses. The temporary lighting of one of the bulbs meant that the racer it symbolized was running a race against horses not represented by bulbs. Each horse ran 30 races interspersed among the contests of the others. For each race, *S* predicted whether the horse represented by the lighted bulb would win or lose. He was then informed of the outcome of that race before the next hypothetical contest was signaled. The horse symbolized by the center light won a predetermined 80% of its races and the other six horses won 20% of theirs.

The results suggested that the strong tendency to bet "win" which had become associated with the 80% winner generalized to neighboring contestants since the 20% winners closer to the 80% winner evoked significantly more win bets than the more distant 20% winners.

The present experiment was conducted to determine whether a similar generalization-like effect would be obtained if all stimuli were presented at the *same* location and differed only in some characteristic independent of spatiality. The characteristic selected was grayness.

Procedure

The spatially undifferentiated stimuli were four silhouettes of a running horse ridden by a jockey. They differed from each other only in degree of grayness. These silhouettes were on 2 × 2-in. slides and were projected one after another onto a white cardboard screen in a black-painted, minimally illuminated room. The approximate size of the racers can be estimated from the fact that a rectangle 19 in. long and 14 in. high would enclose the projected figures. As measured by a Macbeth-Ansco Color Densitometer, Model 12, the densities of the silhouettes on the slides were 2.09 (almost black), .95 (a dark gray), .51 (a medium gray), and .30 (a light gray). These densities are roughly equidistant on a log scale. The background density for all slides was .14.

The brief projection of the silhouette of a racer signified a race. After it was over, *S* bet whether the horse represented by the projected silhouette won or lost and then *E* told *S* the outcome of the race. The *S*s were also told they could maximize the accuracy of their predictions by noting how often each horse won.

Each horse raced 40 times. One of eight prearranged serial sequences of slides was used for each *S*. Each sequence was composed of eight blocks of 20 slide presentations, but for each sequence the eight blocks were in a different random order. The succession of slides within a block was random except that (*a*) each slide appeared five times, (*b*) none appeared more than

[1] This article is based on a Master's thesis submitted to the Graduate College of the State University of Iowa. The writer is indebted to Judson S. Brown for his direction of the research and for his assistance in the preparation of this manuscript.

[2] Now at Skidmore College.

twice consecutively, and (*c*) each one appeared either two or three times in the first half of the block.

Within each block, one horse won four of its five races (80%), and the other three racers each won twice (40%). For one group of *Ss*, Group B, the 80% winner was the blackest horse. For a second group of *Ss*, Group LG, the lightest gray horse won 80% of its races. For each group the particular races won by each horse were randomly determined.

Since the betting behavior of primary interest was that which took place after *S* had had some experience with the actual win-lose frequencies associated with the racers, the first 10 races of each horse were delegated an informative function only and were ignored in obtaining the criterion measures. The criterion measure was the number of win bets placed on each horse during the last 30 of its 40 races.

The *Ss*, 64 students enrolled in undergraduate psychology courses at the State University of Iowa, were randomly divided between Groups B and LG. Each of the eight serial sequences of slides was used with four randomly selected *Ss* from each of these two groups.

Prior to the betting session *Ss* were required to identify correctly the racer shown on each of 12 additional slide presentations. The four racers appeared three times each somewhere within this series. If an *S* made any mistakes on his first 12 trials, a different order of the slides was used to give him a second (and if necessary, a third) chance to satisfy this criterion of successful discrimination. In this period as well as in the betting session each slide presentation lasted 2 sec. and the average interpresentation interval was approximately 10 sec. Four *Ss* who failed in this task were discarded along with two others who were uncooperative. Thirty-seven *Ss* met the criterion on their first try, 20 on their second, and 7 on their third.

Results and Discussion

The percentages of win bets placed on each horse during its last 30 races are plotted in Fig. 1. On the abscissa are shown the densities of the silhouettes of the racers to which the percentages correspond. When the data for Groups B and LG were combined, the mean percentage of win responses for the two 80% winners was 80.1. The mean percentages for the horses whose darkness values differed increasingly from those of the 80% winners were 54.7, 47.8, and 40.2, respectively. Clearly, the 80% winner elicited more win responses than the other three racers, and, as darkness differences between the 80% winner and the other racers increased, the proportion of win responses elicited by them declined. The results are thus comparable to those of Brown, Clarke, and Stein (**2**).

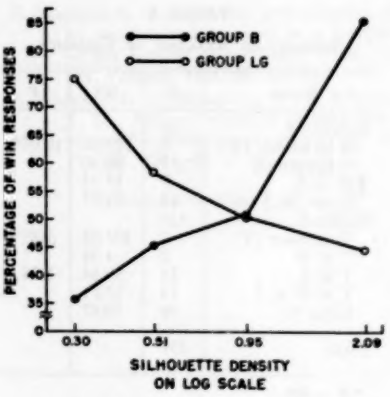

Fig. 1. Mean percentages of win responses made by Groups B and LG during the last 30 presentations of each stimulus.

The finding that the 80% racers elicited more win bets than the others is not, of course, indicative of generalization. The generalization data consist of the betting differences attending the 40% racers. Accordingly, the win frequencies for the 40% winners only were evaluated by the analysis of variance (Lindquist Type III, [**7**]) summarized in Table 1. The differences in the mean number of win responses elicited by these racers, when grouped according to the number of steps in density removing them from the favored racer, were significant ($P < .001$). Further analysis showed that this treatment effect was significant in both Group B ($P < .01$) and Group LG ($P < .02$).

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Between-*S*	63		
80% winner (W)	1	218.88	10.05*
Sequence (S)	7	20.20	
W × S	7	17.92	
Error (b)	48	21.77	
Within-*S*	128		
Treatment (T)	2	302.08	9.85**
T × W	2	4.94	
T × S	14	39.14	1.28
T × W × S	14	17.71	
Error (w)	96	30.67	
Total	191		

* $P = .005$.
** $P = .001$.

During the initial portion of the betting session *S*s displayed a preference to bet "win" when shown the black horse. This preference might have contributed to the significant difference ($P < .005$) between Groups B and LG. No initial preference to bet "win" attended the light gray horse. The initial betting-behavior data also revealed that in Group B, no gradient or monotonic trend among the three 40% winners was present at the start of the experiment. Group LG exhibited a slight initial gradient among the 40% racers, but its slope was opposite to that of the final gradient shown in Fig. 1. It may be concluded that neither of the two final gradients can be accounted for solely in terms of initial response tendencies. A more tenable conclusion seems to be that the strong tendencies to say "win" to the black horse in Group B and to say "win" to the lightest gray horse in Group LG generalized to the 40% racers as a function of their distance on the grayness scale from the favored racer.

Although information was furnished during the experiment that all six 40% winners won 40% of the time, five of these six mean percentages exceeded 40% (see Fig. 1). Generalization of the dominant win response associated with the 80% winner could easily account for this finding. However, it can be explained more simply in terms of initial response tendencies since the six percentages elicited by these racers on their very first race were greater than 40%.

SUMMARY

Previous research indicates that generalization-like effects can be obtained in a betting situation involving stimuli identical except for their spatial location. The present experiment was conducted to determine whether a comparable effect would be obtained with spatially undifferentiated but distinctive cues.

College students were asked to predict the outcomes of a series of simulated horse races. The stimuli were four silhouettes of a racer differing only in degree of grayness. A race was represented by a brief projection of one of the silhouettes on a screen. The *S*s guessed after each race whether the horse represented by the projected silhouette won or lost. After each bet *E* told *S* the outcome of that race. Each horse raced 40 times. Three horses won 40% of their races, and a fourth horse won 80%. For one group of *S*s, the 80% winner was the darkest gray horse, while for the other group, it was the lightest gray racer.

The mean percentages of win bets placed on each horse during its last 30 races were computed separately for the two groups. In each group a direct monotonic relationship was obtained between the number of win responses evoked by a 40% winner and its closeness in darkness value to the 80% winner. These empirical gradients of response frequency were interpreted as reflecting the presence of generalized reactive tendencies.

REFERENCES

1. BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.
2. BROWN, J. S., CLARKE, F. C., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.
3. GIBSON, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.
4. GRANT, D. A., & SCHILLER, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.
5. GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli

varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.

6. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
8. Margolius, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.
9. Wickens, D. D., Schroder, H. M., & Snide, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

(Received December 2, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

EFFECTS OF PREVIOUS EXPERIENCE AND INFORMATION ON PERFORMANCE ON A WORD-FORMATION PROBLEM [1]

WILLIAM F. BATTIG

University of Virginia

Word formation is a type of verbal problem in which *S*s identify words by naming the letters of the alphabet contained in them, the object being to identify the word by naming as few letters as possible. Performance on this task has been shown to be highly dependent on the specific letters contained in the word (1). Subjects would typically guess letters in a relatively consistent order from word to word, so that words containing primarily letters usually guessed early in the sequence were more easily solved. Despite considerable individual variation, a correlation of .82 was obtained between actual average performance on the various words and that predicted on the basis of the average letter-guessing order for the group. Furthermore, no significant improvement or other change in performance resulted from practice on the task. Since an optimal letter-guessing sequence was rarely if ever adopted, this insensitivity to change with practice is somewhat surprising. Evidently the experimental conditions made it difficult to compare the relative efficiency of various possible letter-guessing sequences, with the result that systematic changes failed to develop.

The present study was undertaken to further examine the nature and stability of letter-guessing sequences in the word-formation problem by introducing conditions into the situation designed to modify these sequences. One such factor was experience with words containing several uncommon letters. One group solved only words consisting primarily of the most commonly used letters of the alphabet, and was compared with two groups given varying numbers of words containing several uncommon letters. Also investigated was the effect of providing information about the relative frequency with which letters of the alphabet are used in the English language.

[1] This research was done in part while the author was a post-doctoral fellow of the National Science Foundation at Stanford University.

METHOD

The word-formation problem has been described previously (1). Initially, *S* is told only how many letters are in the word to be identified. Each time a letter is guessed, *S* is told if this letter is in the word or not, and if so, its position in the word (first letter, second letter, etc.). The object is to identify the word by naming as few letters as possible, the number and order of guesses being recorded by *E*. The original procedure was modified by providing knowledge of results after each word and allowing *S* to guess at the word at any time. If the word was correctly identified before all letters had been named, the reported score was only the number actually guessed up to that point. Any letters not actually guessed were added to the scores prior to analysis to permit comparison with previous results.

Subjects were 96 introductory psychology students at Stanford University, all of whom solved a total of 20 six-letter words. These are presented in Table 1. Two distinct types of words are included, differing in the letters they contain. Words of one set (common) include no more than one of the 13 infrequently guessed letters F G H J K P Q U V W X Y Z. In contrast, each word of the second set (uncommon) contains a minimum of three of these uncommon letters. The first five words (pretest) and last five (posttest) were identical for all *S*s, and all met the criteria of the common set. The middle 10 words (6 through 15) differed for three groups.

Group 0 had no words from the uncommon set, and solved the 10 words from the common set as words 6–15. Group 10 solved the 10 words of the uncommon set, while a third group, Group 5, solved five words from each set as indicated in Table 1. Words of both sets were approximately matched with respect to frequency of usage, as well as number of vowels and vowel-consonant structure. All words were relatively common, occurring at least 11 times per million words according to the Thorndike-Lorge count (6).

Information was varied by providing half of each group with a worksheet on which letters of the alphabet were listed in order of frequency of usage (5). The remaining *S*s used similar worksheets, but the letters were listed in alphabetical order.

Half of each subgroup was scored in terms of time to solution as in the previous study, while the other half was told that time was unimportant, and no record of time was kept.

To summarize, a $3 \times 2 \times 2$ factorial design was used with 12 groups of 8 *S*s each. Three levels of experience on words with uncommon letters were included, along with two types of frequency information, and two timing conditions.

After the experiment, all *S*s were questioned to determine whether the differences between the common and uncommon sets of words had been noted. In the course of questioning, the possibility of differences in letter structure was suggested, and *S*s were permitted to actually look at the words on their worksheets.

Results

In addition to the basic *trials* score (number of letters guessed for each word), a *recognition* score was obtained, consisting of the number of letters in the word which had not actually been guessed when correct identification was made. Analysis of letter-guessing sequences was based on a *letter preference score*, computed by assigning a score of 25 to the first letter guessed for each word, 24 to the second, etc., so that if all letters were named before a word was solved, the last letter had a score of zero. Any letter not guessed for a particular word was also given a score of zero. Therefore, the magnitude of the letter preference score reflects the likelihood of the letter being guessed early in the sequence.

Group means for each measure are presented in Table 2. Application of Bartlett's test (2) showed no significant departure from homogeneity of variance. Therefore, all measures were analyzed by means of a triple

TABLE 1

Mean Trials Score for Each Word Used

Pretest			Common Set			Uncommon Set		
No.	Word	Mean Trials	No.	Word	Mean Trials	No.	Word	Mean Trials
1	Dragon	13.00	6	Source	12.94	6*	Upward	16.64
2	Active	13.88	7	Maiden	12.13	7*	Voyage	19.22
3	Modest	12.03	8*	Costly	14.11	8	Punish	14.44
4	Barley	15.14	9*	Branch	14.77	9	Squeak	16.47
5	Listen	11.76	10	Notice	9.88	10*	Joyful	18.81
			11	Shield	12.25	11*	Plague	14.59
Posttest			12*	Impart	11.91	12	Worthy	16.75
			13	Melody	12.72	13*	Quiver	19.06
			14*	Fasten	10.80	14	Knight	16.28
			15*	Tailor	9.84	15	Figure	14.22
16	Market	11.24						
17	Signal	12.41						
18	Credit	11.10						
19	Throne	11.48						
20	Prison	10.96						

* Words given to Group 5.

TABLE 2

MEAN PERFORMANCE MEASURES

Factor	Trials		Recognition		LPS Proportion
	Pretest	Posttest	Pretest	Posttest	Posttest
Experience					
Group 0	13.34	10.87	1.04	1.39	0.141
Group 5	13.46	11.73	1.13	1.34	0.169
Group 10	12.68	11.71	1.23	1.29	0.177
Information					
Frequency	12.61	11.03	1.12	1.37	0.154
Nonfreq.	13.71	11.84	1.15	1.30	0.177
Timing					
Timed	13.25	11.59	1.05	1.25	0.166
Untimed	13.07	11.28	1.22	1.42	0.165

classification analysis of variance, which included the variables of previous experience, information, timing, and all interactions between them. Since in no case did an interaction approach significance, the results for the three main variables will be presented and discussed separately.

Previous experience.—Of primary interest with respect to the experience variable is a comparison of Groups 0, 5, and 10 on the five posttest words. If experience with words of the uncommon set increases the tendency to guess uncommon letters, poorer posttest performance should result, where these letters rarely occur. Although Table 2 shows that the Group 0 posttest trials mean was below Groups 5 and 10, the difference was not significant ($F = 2.03$, 2 and 84 *df*). However, since pretest words were identical for all groups, it was possible to adjust for differences in initial ability, thereby eliminating posttest variability attributable to such differences. The difference between the pretest and posttest trials scores was therefore determined for each *S*, and subjected to an identical analysis. These differed significantly for the three groups ($F = 4.19$, 2 and 84 *df*), closely approximating a straight line relationship with progressively smaller pre-posttest differences from Group 0 to Group 10. Application of Grant's extended trend test (**3**) showed that over 99% of the between-group variance was accounted for by the linear component of the function ($F = 8.37$, 1 and 84 *df*).

Analysis of the recognition score yielded similar results. Although Group 0 was slightly superior to the other groups on the posttest, this was nonsignificant ($F < 1$). However, analysis of the pre-posttest difference scores showed a significant group difference ($F = 3.82$, 2 and 84 *df*). This relationship is also linear as shown by the highly significant linear component emerging from the trend analysis ($F = 7.65$, 1 and 84 *df*). Experience with words from the uncommon set not only adversely affected the posttest trials score, but also led to an impairment in ability to recognize the posttest words when only a few letters were known.

The letter preference scores were used to compare groups with respect to letter-guessing sequences. When

each letter was considered individually, no significant difference in pattern of letter preference scores was found, and marked individual variation was shown. However, this is an insensitive test of the effect of experience with words of the uncommon set on letter-guessing sequences. To provide a better test, the proportion of the total letter preference scores accounted for by the 13 infrequently guessed letters stressed in the uncommon set (*LPS proportion*) was determined for each *S* on the posttest words. These means are presented in the last column of Table 2, and show an increased tendency to guess uncommon letters with increased experience with words containing these letters. These differences are significant ($F = 4.84$, 2 and 84 *df*), and the linearity of the relationship is shown by the highly significant linear component ($F = 9.52$, 1 and 84 *df*) which again accounts for all but a small fraction of the between-group variance. It is apparent that group differences in performance are associated with differential tendencies to guess infrequently used letters, resulting from experience with words containing these letters.

Information.—Subjects given information about the relative frequency with which letters are used in the English language displayed significantly superior performance on both pretest ($F = 7.16$) and posttest ($F = 4.10$) words, both with 1 and 84 *df*. No significant effect of information was found in the pre-posttest difference trial scores, indicating that the effect is present initially, and is not appreciably changed by practice. The information variable did not significantly affect performance on Words 6–15 ($F = 1.19$), and no significant difference was obtained for either pretest or posttest recognition measures.

The letter preference scores showed a significant effect of frequency information. Analysis of variance of the letter preference scores for individual letters revealed a significant information by letters interaction ($F = 4.88$, 25 and 2100 *df*). An adaptation of the trend test (**3**) was used to further analyze this interaction, by comparing trends for the two information conditions when ordered both alphabetically and in terms of frequency of usage. The group given frequency information showed a significantly lower slope for alphabetical order ($F = 4.16$, 1 and 2100 *df*), and a higher slope for frequency ordering ($F = 25.82$), indicating a greater correspondence of letter-guessing sequences to the order of frequency of usage for *S*s given this information, and a closer correspondence to alphabetical order for *S*s not given this information. However, the difference between the two information conditions in LPS proportion scores is of only borderline significance ($F = 3.92$, 3.96 required for .05 level with 1 and 84 *df*). This contrasts with the results with the experience variable, but provides evidence for the independence of the two variables in their effects on letter-guessing sequences.

Timing.—Whether or not *S*s were timed had little if any effect on performance. Untimed *S*s were slightly superior to timed *S*s on most measures, but this was significant in only one case. The untimed group had a higher recognition score, significant on the pretest words ($F = 8.29$, 1 and 84 *df*), but not on the posttest ($F = 2.21$). This suggests that in the initial stages, timing leads to reduced success in identifying the word when more than one letter is still unknown.

Word differences.—Sizeable differences between the common and uncommon sets were found, as shown by the mean trials scores presented in Table 1. With one exception, all words of the common set have lower means than any word of the uncommon set, while only one of the pretest and posttest words overlaps with the uncommon set. In addition, highly significant differences ($F = 89.19$) were found between the mean trials scores for Groups 0, 5, and 10 on Words 6–15. Recognition scores were also found to differ significantly, being lower for words of the uncommon set ($F = 4.58$, 2 and 84 *df*).

The post-experimental questioning showed little if any recognition of the difference between the common and uncommon sets. Only 4 of the 32 *Ss* in Group 10, where these differences were most apparent, gave any indication of having noticed such a difference. These *Ss* showed a slightly greater tendency to guess uncommon letters on the posttest, as measured by the LPS proportion scores, but not enough to account for the observed effects.

Discussion

It can be concluded from the present study that performance on the word-formation problem is affected both by experience with words containing uncommon letters and information about frequency of letter usage. Furthermore, both variables produced changes in the letter-guessing order. This supports the conclusion of the previous study (1) that the primary determinant of word-formation performance is the degree to which letters in the word are guessed early in the sequence. However, these sequences were shown to be modifiable in the present study. The resulting performance changes are small in magnitude, but the variation in experience given Groups 0, 5, and 10 was limited, and it is likely that more extensive variation would lead to more sizeable effects.

The effect of the experience variable can be readily interpreted within a reinforcement framework. Solution of words containing uncommon letters presumably reinforces the guessing of such letters, with a resultant increase in the probability of guessing uncommon letters on subsequent words. The process involved here appears similar to that involved in studies of "verbal conditioning," as is further suggested by the inability of most *Ss* to distinguish between the common and uncommon sets of words. This corresponds to the "learning without awareness" characteristic of many verbal conditioning studies (e.g., 4). Although the post-experimental questioning was relatively direct, the role of awareness was not a consideration in this study, and the adequacy of the measure of awareness is questionable.

The significant effect of experience on the recognition scores indicates that shifts in letter-guessing sequences are not the only factor leading to performance changes. This is a measure of ability to recognize words when several letters are still unknown, which apparently depends on what the letters are. The finding that recognition scores are lower for the uncommon set of words indicates that recognition is easier if the missing letters are common ones.

Summary

Performance on a word-formation problem, in which *Ss* identify words by naming the letters of the alphabet contained in them, was investigated under conditions designed to modify the order in which the letters were guessed. Increased experience with words containing uncommon letters led to an increased tendency to guess such letters, and a subsequent decrement in performance on words with common letters. Information about the relative frequency of letter usage led to superior performance with letter-guessing sequences corresponding more closely to frequency order. The results support previous evidence that performance on this task is highly dependent on the letters constituting the

word, and their relationship to *S*'s typical letter-guessing sequence.

References

1. Battig, W. F. Some factors affecting performance on a word-formation problem. *J. exp. Psychol.*, 1957, **54**, 96–104.
2. Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.
3. Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.
4. Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
5. Pratt, F. Secret and urgent. Garden City, N. Y.: Blue Ribbon Books, 1942.
6. Thorndike, E. L., & Lorge, I. The teacher's wordbook of 30,000 words. New York: Teachers Coll., Columbia Univer., 1944.

(Received December 3, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

TIME SHARING AS AN INDEX OF AUTOMATIZATION [1,2]

HARRY P. BAHRICK AND CAROLYN SHELLY
Ohio Wesleyan University *Ohio State University*

It has long been held (**6**, Ch. 4) that *Ss* can take advantage of redundant elements in sequential tasks by changing their method of responding. In such tasks proprioceptive stimulation associated with one response may come to elicit the next response in the sequence so that *S* becomes less dependent upon exteroceptive cues. This presumed gradual change from exteroceptive to proprioceptive control during prolonged practice of repetitive tasks has been labeled "automatization" (**9**, p. 27). Frequently such a sequential conditioning process has been assumed (**10**), but has rarely been investigated in the laboratory. No operations are known which would define the process, and none of the conventionally used indicants of learning will allow us to infer it.

Recently, particularly since information theory has made it possible to quantify the degree of redundancy in sequential tasks, human ability to deal with redundant aspects of the environment has become a matter of considerable research interest. Redundancy of perceptual, motor, and language activities has been measured (**1, 8**), and the ability of *Ss* to detect the presence of redundancy has also been investigated (**3**).

The present study attempts to relate the degree of redundancy of a task to the degree of automatization, or proprioceptive control, that can be reached during extended practice, and it attempts to infer the presence of this type of control through performance in a time-sharing situation. Specifically, *Ss* are given prolonged training on one of four versions of a serial reaction task to light stimuli. The task versions differ in degree of redundancy of the stimulus sequence. At three stages of training, a second, auditory reaction task is added and *Ss* are required to perform the two tasks concurrently. Interference effects due to the time-sharing requirement are determined and are related to the redundancy of the visual task. The usefulness of a time-sharing situation in this context was established in an earlier study (**2**) in which combined performance on two tasks was observed for a random and a repetitive version of one of the tasks. However, redundancy of the tasks was not quantified and no functional relations between the variables could be established.

METHOD

Visual Task

The principal pieces of apparatus were a stimulus panel, a response panel, and a programming unit. The stimulus panel consisted of four 6-v. lights spaced horizontally at a distance of 1.75 in., with a fifth, knowledge-of-results light, placed midway between the second and third light and 3 in. above them. The panel was mounted at eye level at a distance of 30 in. facing the seated *S*. The response panel consisted of four telegraph keys spaced at .75-in. intervals, mounted horizontally in a convenient position for four fingers of *S*'s right hand, not including the thumb. The four response keys corresponded to the four stimulus lights on a simple spatial basis, so that the key at the extreme left

[1] This research was supported by a grant from the National Science Foundation to Ohio Wesleyan University.

[2] The writers wish to thank G. E. Briggs and P. M. Fitts for many constructive criticisms.

was appropriate to the light at the extreme left, and so forth.

The programming unit consisted of a rotary stepping switch with 103 positions, and seven poles for each position. Each position of the rotary stepping switch could be connected with any of the four lights on the stimulus panel by means of leads provided with alligator clips. Thus, it was possible to program any sequence among the four stimulus lights, which would not repeat until 103 stimuli had occurred. The stepping switch rotated continuously at a rate of 1 rpm, so that each stimulus was present for about .58 sec. If one of the response keys was depressed during an interval in which the corresponding stimulus light was present, a hit was registered on one of four corresponding counters, and the knowledge-of-results light was lit indicating a hit. Four versions of this task were used:

Repetitive version (*RP*).—The sequence of lights presented was repeated in an identical pattern after every fourth stimulus. Numbering the lights from left to right, 1 through 4, the pattern used was 1, 3, 2, 4, 1, 3, 2, 4, ... A trial consisted of 100 stimuli, i.e., the sequence repeated 25 times. The *S*s were instructed to press the key corresponding to the light which appeared as quickly as possible with the appropriate finger of the right hand. The *S*s were told that the knowledge-of-results light on the stimulus panel would appear each time they registered a hit.

High redundant version (*HR*).—All four of the lights occurred with equal frequency (25 times), but 4 digrams (1–3, 3–2, 2–4, and 4–1) each occurred 16 times, and the remaining 12 digrams each occurred 2 or 3 times. Within these constraints, trigram frequencies were randomized.

Low redundant version (*LR*).—All four of the lights occurred with equal frequency (25 times). The 4 digrams (1–3, 3–2, 2–4, and 4–1) each occurred 10 times and the remaining 12 digrams each occurred 4 or 5 times. Within these constraints, trigram frequencies were randomized.

Random version (*RA*).—Monogram and digram frequencies were arranged to occur with approximately equal frequency so that each light appeared 25 times and each digram occurred 6 or 7 times. Within these constraints, the sequence was randomized.

The maximum entropy per symbol for all four task versions was 2 bits, whereas the actual entropy was 0, 1.74, 1.96, and 2 bits; thus, the relative redundancy ($R = 1 - H_{actual}/H_{maximum}$) values were 100%, 13%, 2%, and 0% for the *RP*, *HR*, *LR*, and *RA* task versions, respectively.

Auditory Task

The principal pieces of apparatus were a 78-rpm phonograph, an amplifier and noise generator unit, earphones, a voice key, an Esterline-Angus recorder, and a response panel consisting of five telegraph keys mounted in a convenient position for *S*'s left hand. At irregularly spaced intervals ranging between .5 sec. and 4 sec., *S* heard numbers over the earphones. The numbers 1 through 5 were used. The sequence was randomized within the constraint that each number occurred three times for a total of 15 stimuli presented over a period of 46 sec. Four different random sequences were recorded on phonograph records. The intervals between stimuli were filled with white noise produced by an electronic noise generator. The noise was conducted through the earphones and was interrupted whenever a number stimulus occurred. The interruption was accomplished by means of a voice key activated by the amplified sound of the number from the phonograph record. The voice key was also used to mark the occurrence of the stimulus on the Esterline-Angus recorder. The *S*'s task consisted of responding to each number read over the earphones by depressing with the fingers of the left hand as quickly as possible the telegraph key which corresponded to that number. The keys were numbered 1 through 5 from left to right. Separate channels of the Esterline-Angus recorder were used to mark the occurrence of stimuli, and to record responses by each of the five fingers. This made it possible to score reaction time as well as errors.

Subjects

Forty male and female student volunteers from elementary and experimental psychology classes at Ohio Wesleyan University were divided by means of a table of random numbers into four groups of 10 *S*s each, and each group was assigned to one of the four visual task versions.

Procedure

During the first experimental session all *S*s received 10 trials on the version of the visual task appropriate to their group. Each trial consisted of 100 stimuli, and after each trial *S*s were informed of the total number of hits scored. A rest period of 1 min. occurred between trials. The second session (Trials 11–20) was conducted exactly like the first one. Before the third session all *S*s were instructed in the auditory task and were told that they would be scored on both the visual and auditory tasks, which were to be considered equally important. The third session consisted of four time-shared trials on the visual and auditory tasks. The particular sequence of auditory stimuli was changed on each trial by a change of phonograph records. Within each trial the two tasks were synchronized in such a way that the first auditory stimulus occurred

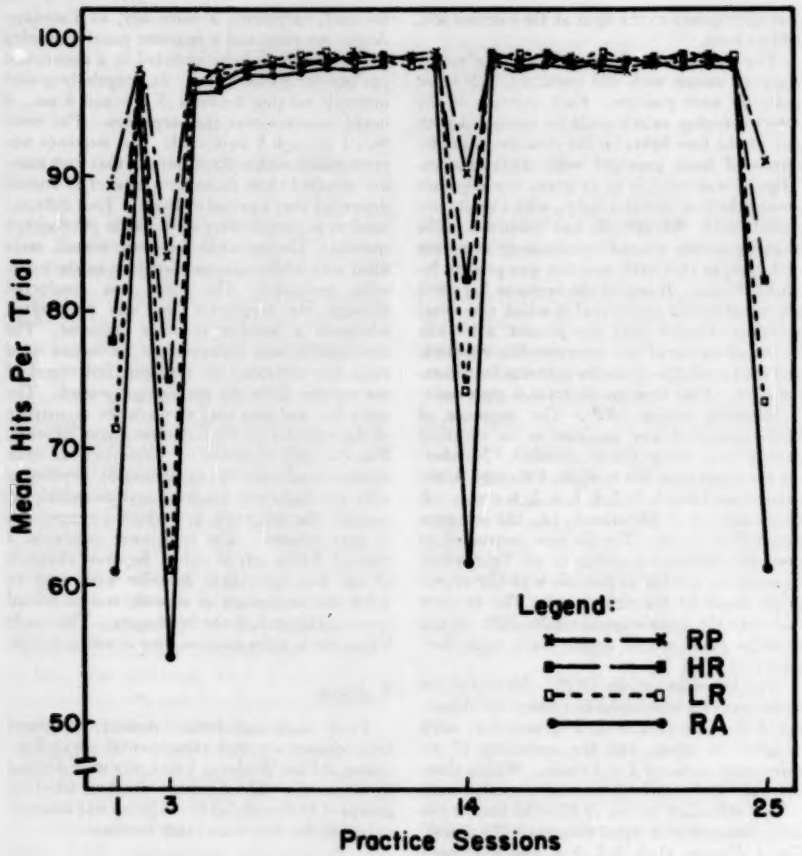

FIG. 1. Performance on the visual task.

after 7 sec. of performance on the visual task. Stimuli of both tasks continued concurrently for 46 sec., and during the last 6 sec. only visual stimuli occurred.

After each trial of time-shared performance *Ss* were informed of the total number of hits scored on the visual task and of the number of correct responses on the auditory task. However, their reaction time to the auditory stimuli was not reported. In all, *Ss* received 25 training sessions. The visual task was used alone on Sessions 1–2, 4–13, and 15–24 with 10 trials per session. Tests under the time-sharing condition were administered on Sessions 3, 14, and 25, with four trials per session. These 25 training sessions were given over a period of six weeks, with no more than one session per day and intervals no larger than three days between sessions.

RESULTS

Figure 1 shows the performance of the groups on the visual task. It can be seen that all four groups reached a high proficiency level within 20 trials, and during the remainder of the experiment none of the group means fell below a level of 97 hits out of a possible 100 when the visual task was presented alone. No significant

TABLE 1

ANALYSIS OF VARIANCE FOR THE VISUAL TASK DURING TIME-SHARED TRIALS

Source of Variance	*df*	*MS*	*F*
Task versions (*TV*)	3	4622.23	43.64**
Ss within groups (*Ss*)	36	105.10	
Training sessions (*TS*)	2	744.29	70.15**
TV × *TS*	6	106.59	10.05**
Pooled *Ss* × *TS*	72	10.61	
Total	119		

** $P < .01$.

differences among the groups are evident on trials where the visual task is practiced alone except at the very beginning of training. An analysis of variance for repeated measures was performed to test for group differences on Sessions 2, 13, and 24 (the sessions just prior to the time-shared trials). The analysis indicates no significant differences ($P > .05$) between groups, between training levels, or between combinations of groups and training levels.

Large and highly significant group differences occurred in the visual task, however, on those trials during which time sharing was required. Table 1 shows the results of a comparable analysis applied to the data of Sessions 3, 14, and 25. It is apparent that the introduction of the auditory task caused interference effects which varied inversely with the redundancy of the visual task. The analysis further indicates that continued practice on the visual task reduced susceptibility to interference. However, inspection of Fig. 1 shows that interference effects did not diminish greatly beyond Session 14.

Performance on the auditory task was scored for cumulative reaction time, correct responses, and errors. Reaction time was cumulated for the first response to each stimulus regardless of whether the response was correct or not. If several responses were made to the same stimulus, each was scored as correct or error. If no response occurred, a maximum reaction time of 1 sec. was assumed and an error was scored. Table 2 shows mean performance per trial for Sessions 3, 14, and 25. Little improvement occurred, and analysis of variance for repeated measures, carried on separately for reaction time, errors,

TABLE 2

CORRECT RESPONSE (*CR*), ERRORS (*ER*), AND CUMULATIVE REACTION TIME (*RT*) PER TRIAL FOR THE AUDITORY TASK

Practice Session		Groups											
		RP			*HR*			*LR*			*RA*		
		CR	*ER*	*RT*	*CR*	*ER*	*RT*	*CR*	*ER*	*RT*	*CR*	*ER*	*RT*
3	*M*	13.5	7.2	8.2	14.1	8.4	7.4	14.0	5.7	6.8	13.5	9.0	7.0
	SD	1.5	3.8	2.4	1.3	4.8	2.7	1.3	3.9	2.9	1.2	4.2	3.6
14	*M*	14.1	5.6	6.6	13.4	7.2	6.9	14.4	7.0	5.6	13.7	6.8	6.2
	SD	1.1	3.8	2.8	1.4	4.0	3.0	1.0	3.1	2.8	1.4	3.6	2.9
25	*M*	14.3	6.3	7.0	14.2	7.2	7.3	14.1	7.4	6.8	14.2	7.2	6.3
	SD	1.3	4.3	3.0	1.1	4.1	3.2	1.3	3.8	3.1	1.2	4.0	2.8

and correct responses, indicates no significant differences ($P > .05$) between groups or between sessions for any of the three scores.

Discussion

It is apparent that time sharing reveals aspects of learning of the visual task which cannot be observed when the visual task is performed alone. The interference effects produced by time sharing are an inverse function of redundancy of the visual task, and these effects can serve as an index of the degree of automatization, or the extent to which *S* is able to make use of redundancy and become independent of exteroceptive stimuli.

Figure 2 shows the relation between redundancy and performance decrement due to time sharing. The data are for Session 25. It can be seen that *S*s are able to make effective use of a relatively small amount of redundancy and that much larger amounts of redundancy produce comparatively little additional protection against interference. The nature of the function in Fig. 2 is determined in part by the sensitivity of the redundancy index to differential probability of stimuli. Fisher (**4**, p. 221) shows in an analogous function that large probability differentials may have relatively minor effects upon entropy. It is apparent, however, that *S*s can use such

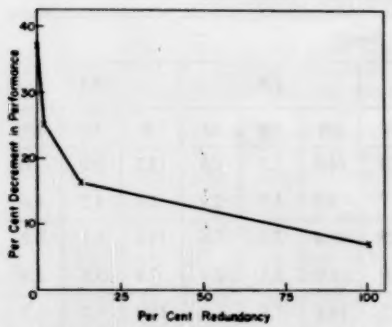

Fig. 2. Performance decrement due to time sharing as a function of task redundancy.

differentials and that time-sharing situations can produce a sensitive index of this aspect of learning. The function in Fig. 2 may also vary with the informational content of the stimuli of both tasks. The relatively low informational content of stimuli in the present task may increase *S*'s ability to take advantage of small amounts of redundancy.

It is not possible to arrive at an accurate estimate of the information transmitted for the various versions of the visual task during the time-sharing trials. This would require knowledge of the error distribution which is not available. If it is assumed that errors are randomly distributed, then the information transmitted is approximately 2.4 bits/sec. for both the *LR* and *HR* task versions and only about 2.0 bits/sec. for the *RA* version. Since errors are usually not randomly distributed, the above are probably underestimates. The information transmitted for the auditory task was computed, and data for all three stages of practice were combined, since mean performance did not improve significantly. The estimates ranged from .6 to .7 bits/sec. for the four groups. The combined information transmitted on both tasks during the time-shared trials probably did not exceed 3.1 bits/sec. for any group. This is somewhat lower than has been reported in other situations (**5**). However, the combined informational content of stimuli from both tasks did not exceed 4.1 bits/sec. for any group. The capacity to transmit information may also be reduced because of the concurrent use of two sensory channels. This may be particularly true if similar responses are required to the two types of input, as was the case here. Substantial interference effects might be expected under such circumstances, and the data on both tasks suggest this. Results obtained by Klemmer (**7**), in another study of time sharing between auditory and visual channels, tend to confirm this. Klemmer, who also used key-pressing responses for both tasks, found that forcing *S* to alternate regularly between the tasks at

a rapid rate lowered his over-all performance sharply.

Summary

Concurrent performance of a visual and an auditory serial reaction task was measured at three stages of practice of the visual task, and for four versions of the visual task differing in degree of redundancy of the stimulus sequence. The degree of redundancy of the visual task had no effect upon performance when that task was practiced alone, all versions of the task being performed at an equally high level of proficiency. It was found that the interference effects due to time sharing of the two tasks varied inversely with the degree of redundancy of the visual task. Relatively small amounts of redundancy produced significant reductions of interference, and differences as a function of redundancy did not diminish with continued training.

The results support the hypothesis that redundancy of stimulus sequences permits a change from exteroceptive control of responses to proprioceptive control, and that performance in time-sharing situations provides a useful index of this process of automatization.

References

1. Aborn, M., & Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.
2. Bahrick, H. P., Noble, M. E., & Fitts, P. M. Extra-task performance as a measure of learning a primary task. *J. exp. Psychol.*, 1954, **48**, 298–302.
3. Bennett, W. F., Fitts, P. M., & Noble, M. E. The learning of sequential dependencies. *J. exp. Psychol.*, 1954, **48**, 303–312.
4. Fisher, R. A. *The design of experiments.* Edinburgh: Oliver and Boyd, 1949.
5. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47**, 381–391.
6. James, W. *The principles of psychology.* New York: Holt, 1890.
7. Klemmer, E. T. Time sharing between auditory and visual channels. In G. Finch and F. Cameron (Eds.), Symposium on human engineering, personnel, and training research. *USAF, ARDC Tech. Rep.*, 1956, No. 56-8. Pp. 199–203.
8. Senders, V. L., & Cohen, J. The effects of sequential dependencies on instrument reading performance. In H. Quastler (Ed.), *Information theory in psychology: Problems and methods.* Glencoe, Ill.: Free Press, 1955. Pp. 282–289.
9. Warren, H. C. *Dictionary of psychology.* Cambridge, Mass.: Houghton Mifflin, 1934.
10. Watson, J. B. *Behaviorism.* (Rev. ed.) New York: Norton, 1930.

(Received December 13, 1957)

Journal of Experimental Psychology
Vol. 56, No. 3, 1958

THE TEMPORAL FACTOR IN REWARD AND NONREWARD OF MONKEYS[1]

R. W. LEARY

University of Oregon

Although investigations (**2, 8**) of the discriminative behavior of monkeys have disclosed that reward and nonreward are associated with distinctly different acquisition functions, Harlow and Hicks (**3**) have suggested that the difference is one of performance and not one of learning. Their interpretation utilizes the concept of response-shift which can be defined as a greater tendency to vary choice following a rewarded selection than following a nonrewarded selection, and which seems to be regarded as an expression of an interfering motivational factor such as curiosity or, possibly, reactive inhibition (**2, 3**). In advocating a uniprocess theory they posit that reward and nonreward should yield identical intertrial delay functions.

Previous work of the present author (**4**) has shown that an unchosen negative object which accompanies a chosen positive object for several trials gains in attractiveness. This finding appears relevant to the response-shift phenomenon originally described by Harlow (**2**) in paired discrimination but is not applicable to the analogous effect discovered by Moss and Harlow (**8**), which is that when an object is presented alone for one or more trials and is then immediately paired with a new object for several trials *S*s make fewer errors if the original object is consistently incorrect (nonrewarded) than if it is consistently correct (rewarded).[2] A modicum of evidence has been found that there are temporal changes in object attractiveness which may be functions of reward treatment (**4, 6**).

The purpose of the present study was to explore the Moss-Harlow effect by use of a serial technique (**5**) which allowed concurrent experience of several conditions. This technique increases the time between successive trials with the same object (or pairs) to about 2 min., compared with perhaps 10 sec. in ordinary discrimination, and also imposes intervening trials with other objects (or pairs). In addition to reward and nonreward conditions, an alternate-value condition was used in order to test the effect of equal reward and nonreward experience of the same object. Discrimination performance was secured both at a 2-min. interval (which was "immediate" in the context of the experiment) and after a delay of 24 hours.

METHOD

The general method of serial presentation has been described elsewhere (**5**). The Wisconsin General Test Apparatus (**1**) and a large set of randomly paired multidimensional objects were used.

Nine test-wise rhesus monkeys learned lists of nine single objects. In every list the objects were randomly assigned to three mutually exclusive reinforcement conditions, three objects to

[1] This research was supported in part by a grant (M-1664) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] The term *response-shift* was originally used to designate a statistically described phenomenon in paired discrimination learning but was also directly identified with the Moss-Harlow effect (**2**). It is regrettable that the term has now been adumbrated by identifying it with several effects (**2, 7, 8**) which are phenomenally different and may lack causal equivalence.

each condition. These conditions were as follows: Cond. R in which objects were consistently rewarded, Cond. NR in which objects were consistently nonrewarded, and Cond. AR in which reward and nonreward of each object were alternated during the training phase. Different nine-object lists were selected at each session for each *S*. Eight runs through each list constituted the training phase, no discrimination being required. A 20-sec. rest was interposed between successive runs.

The discrimination phase followed the training phase immediately in four sessions and after 24 hr. in four sessions. The sequence of the immediate and delayed sessions was balanced according to an ABBABBA schema. Discrimination tests were conducted by pairing the previously and singly presented objects of the training phase each with a new object and by presenting the nine pairs thus constituted on two successive runs. Only one object in each pair was rewarded: in both Cond. R and AR this was the previously presented object and in Cond. NR this was the new object.

Results

An analysis of variance performed on error scores revealed that reinforcement condition was significant ($P < .001$), that time of test was not significant, but that their interaction was significant ($P < .001$). However, since error scores were not normally distributed, low scores being relatively frequent, further analysis was carried out with Wilcoxon's nonparametric test for matched pairs.

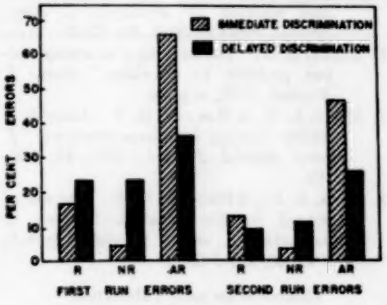

FIG. 1. Errors in immediate and delayed discrimination tests for the three reinforcement conditions.

First-run errors.—The percentages of errors in the three reinforcement conditions are shown in Fig. 1. There was a striking difference in immediate discrimination where the comparison of Cond. R and NR yielded $P = .05$ and where both comparisons involving Cond. AR yielded $P < .01$. At 24 hr. no comparison produced a significant difference. There was below chance (less than 50%) selection of the alternate-value objects in the immediate tests ($P = .02$) but above chance (more than 50%) selection of these objects in the 24-hr. tests ($P = .05$).

Second-run errors.—Scores were generally in the same relative direction as in the first run, as can be seen in Fig. 1, but all comparisons lacked statistical significance except those involving Cond. AR. More errors occurred in this condition than in the other two conditions in both immediate and delayed discrimination tests (no P larger than .02) The negative quality of the alternate value object was clearly discernible in immediate tests where, despite the practice afforded by the first run, choices approximated chance.

Error changes with time.—The loss in accuracy of the first-run choices between immediate and 24-hr. tests was not statistically significant in Cond. R but was significant ($P = .01$) in Cond. NR. A gain in first-run accuracy was found in Cond. AR ($P = .01$). None of the changes in second-run scores were statistically significant.

Discussion

The results indicate that for test-wise rhesus monkeys different learning mechanisms operate for positive and negative objects in discrimination learning. The Moss-Harlow effect (8) was secured in the immediate retention tests: significantly more errors occurred in pairs

containing the positive objects than in those containing the negative objects. The fact that this effect occurs here under conditions involving a relatively long time interval and serial presentation militates against an explanation couched in terms of reactive inhibition and is neutral in respect to explanations based on curiosity or on an assumption of different primary processes. Attenuation of the error difference at the 24-hr. retention test gives an advantage to the latter hypothesis, for there is no immediately apparent reason why investigatory tendency should dissipate while retention remains relatively unaffected. Clearly, the expectation of Harlow and Hicks (3) did not prevail.

The results obtained with the alternate-value objects indicate differential effects for reward and nonreward. These objects which were given partial reinforcement, a condition sometimes associated with rapid positive learning, were chosen less often than chance in immediate tests and more often than chance at 24 hr. This may be interpreted as a decline in the effect of the nonreward experience and/or a gain in the effect of the reward experience. Thus the effects that occurred separately for the rewarded and nonrewarded objects occur together and in nearly additive fashion for the alternate value object. A departure from additivity is suggested in the delayed tests where there was equality of first-run scores in Cond. R and NR and there was a departure from chance selection in Cond. AR.

The present results appear to place a strain on a single explanatory (uniprocess) learning principle. Although hypotheses concerning curiosity and reactive inhibition were not bolstered, no evidence was produced that negated them. Tentatively, one suggestion will be made regarding the phenomena observed. This is that a short-lived negative conditioning effect follows choice of an unrewarded object. Whether this effect is more in the realm of cognition or of emotion is a semantic as well as an experimental problem.

Summary

Nine test-wise rhesus monkeys were given eight runs through lists composed of nine single objects. In each list three objects were rewarded (Cond. R), three were nonrewarded (Cond. NR), and three were alternately rewarded and nonrewarded (Cond. AR). Next, these objects were paired, each with a new object, and were presented for two test runs. In four sessions the paired presentations followed the single object presentations immediately and in four sessions they followed after a 24-hr. delay. In each pair only one object was correct (rewarded); in Cond. R and AR this was the previously presented object, and in Cond. NR this was the new object.

In immediate tests significantly more errors occurred in Cond. R than in Cond. NR and significantly above chance errors occurred in Cond. AR. In the 24-hr. tests Cond. R and NR yielded equal error scores but significantly below chance errors occurred in Cond. AR. The above results pertain to first-run scores; similar trends were found in second-run scores.

References

1. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
2. Harlow, H. F. Analysis of discrimination learning by monkeys. *J. exp. Psychol.*, 1950, **40**, 26–39.
3. Harlow, H. F., & Hicks, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, **64**, 104–109.
4. Leary, R. W. The rewarded, the unrewarded, the chosen, and the unchosen. *Psychol. Rep.*, 1956, **2**, 91–97.
5. Leary, R. W. Analysis of serial-discrimination learning by monkeys. *J. comp. physiol. Psychol.*, 1958, **51**, 82–86.
6. Leary, R. W. The learning of an ambiguous-cue problem by monkeys. *Amer. J. Psychol.*, 1958, in press.
7. Moon, L. E., & Harlow, H. F. Analysis of oddity learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1955, **48**, 188–194.
8. Moss, E. M., & Harlow, H. F. The role of reward in discrimination learning in monkeys. *J. comp. physiol. Psychol.*, 1947, **40**, 333–342.

(Received for early publication January 30, 1958)

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price $1.00

Discounts on quantity orders over 50 copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION

Publications Office, 1333 Sixteenth Street, N.W., Washington 6, D. C.

A *very considerable* proportion of all published articles on perceptual-motor skills and on perception-as-influenced-by-experience are given immediate publication in

PERCEPTUAL AND MOTOR SKILLS

If you are interested in these fields, you and your library should have current subscriptions to PMS ($10.00 per year, with a special rate of $5.00 to individual members of the APA). Prices of back volumes may be obtained on request from

PERCEPTUAL AND MOTOR SKILLS
PSYCHOLOGICAL REPORTS
Box 1441
Missoula, Montana